FIVE THOUSAND WORD SYNOPSIS ON

Identification and molecular characterization of important flowering genes in *Bambusa tulda* Roxb. with major emphasis on photoperiodic pathway and integrator genes

The five thousand word synopsis submitted for requirement of registration for the degree of Doctorate in Philosophy

Submitted by: Smritikana Dutta

PhD Registration no. R-17RS18210090 Dated 14.08.2018

Supervisor: Dr. Malay Das

Plant Genomics Laboratory

Department of Life Sciences

Presidency University, Kolkata

Abstract

Bamboos belong to the subfamily Bambusoideae, family Poaceae and are phylogenetically very close to the reference monocot plant rice (*O. sativa*). They display extensive variations with respect to flowering time, which may be extended up to 120 years. Besides the unusual delay in flowering, another important feature is mass death of all the flowering culms known as monocarpy or semelparity. Therefore, bamboo flowering has an enormous effect on forest ecology as well as economy. Induction of flowering is the most vital developmental events in plants' life. It takes place in a favorable environment, which is decided by a complex regulatory crosstalk at the molecular level. Signals from multiple external and internal sources need to be synchronized to decide the timing of flowering. The two major environmental stimuli that influence flowering are light and temperature. In particular, duration and nature of light plays very important role, based on which plants can be characterized as long day (LD), short day (SD) and day neutral plants (DNP).

The main focus of this study is to understand the role of light on bamboo flowering. Therefore, in this study, four circadian clock genes (*LATE ELONGATED HYPOCOTYL*, *TIMING OF CAB EXPRESSION1*, *ZEITLUPE* and *GIGANTEA*) and two clock integrators (*CONSTANS A, B*) of photoperiodic pathway were identified and sequenced from the tropical bamboo *Bambusa tulda*. Detailed studies on the expression patterns of these genes in different tissues, seasons and diurnal conditions indicate their possible involvement in flowering. Finally, *in planta* functional characterization of *BtFD1* and *BtFD2* genes provide new clues about their possible functional diversification with respect to flowering.

Table of Contents

1. Introduction……..……1-3

2. Objectives (as specified in approved 1000 word)……..……3-4

3. Methodology……...……4-15

3.1. Collection of tissues for isolation of nucleic acids…….4

3.2. Retrieving genomic and protein sequences from different online sequence resources……..……..5

3.3. Isolation of genomic DNA……..……...6

3.4. Designing of primers for gene amplification and expression analysis…..…6

3.5. Amplification of the targeted flowering genes by PCR…….....9

3.6. Calcium competent cells (*E. coli Dh5α*) preparation……....……9

3.7. Gel elution of the amplified PCR products having desired molecular Weight…….……..…..10

3.8. Cloning of the gel eluted PCR products and transformation…….…10

3.9. Colony PCR……10

3.10. Isolation of plasmids and Sanger's sequencing……10

3.11. Prediction of gene models and *in silico* characterization of predicted protein sequences……11

3.12. Phylogenetic analyses……...…….12

3.13. Extraction of total RNA…….12

3.14. Preparation of cDNA libraries……12

3.15. Real time RT-qPCR analyses…….12

3.16. *In silico* prediction of cis-elements and construction of dendrograms..…13

3.17. Gateway cloning and *Agrobacterium* mediated transformation of two *BtFD* gene homologs……...…13

3.18. *In planta* transformation of *BtFD1* and *BtFD2* genes under 35S promoter in *Arabidopsis thaliana* Col-0…….14

3.19. Selection and molecular confirmation of putatively transformed plants...14

3.20. Phenotypic analyses of the transgenic plants…….14

3.21. *In silico* interaction of *BtFD1* and *BtFD2* proteins with the promoter of *AP1* gene……..14

4. Results

Objective 1: Identification and molecular characterization of important circadian clock genes in bamboo……...…….15-38

4.1. *In silico* identification of four important circadian clock genes from all the sequenced genomes of monocotyledonous plants…….……..15

4.2. Identification and molecular characterization of *LHY, TOC1, ZTL and GI* homologs in *B. tulda*……...15

4.3. Phylogenetic relationships of *LHY, TOC1, ZTL and GI* homologs of *B. tulda* with homologs identified from related monocotyledonous species……...17

4.4. Comparative analyses of gene structures and predicted amino acid sequences of *BtLHY*, *BtTOC1*, *BtZTL* and *BtGI* homologs with that of related monocotyledonous species……..19

4.5. Tissue specific gene expression analyses of *BtLHY*, *BtTOC1*, *BtZTL* and *BtGI* in *B. tulda*……...22

4.6. Study of the diurnal expression patterns of *BtLHY, BtTOC1, BtZTL and BtGI* genes in selected flowering and non-flowering tissues…….…….22

4.7. *In silico* identification and comparisons of *cis*-acting elements present in the putative promoter regions of *LHY, TOC1, ZTL* and *GI* genes……24

Objective 2: Identification, sequencing and detailed molecular characterization of important circadian clock integrator gene, *CONSTANS* in *B. tulda* Roxb.

4.8. *In silico* identification of *CONSTANS LIKE* genes from all the Poaceae plants having sequenced genomes……..25

4.9. Isolation and molecular characterization of *CO* homologs from *B. tulda*...26

4.10. *In silico* prediction of gene structures of *BtCOA* and *BtCOB* and comparison with homologs from related monocotyledonous species…….26

4.11. Phylogenetic relationships of *BtCOA* and *BtCOB* with homologous *COL*s identified from related monocotyledonous species…….....28

4.12. Study of gene expression patterns of *BtCO* homologs in different flowering and non-flowering tissues……...29

4.13. Study of the diurnal expression patterns of *BtCOA* and *BtCOB* genes in selected flowering and non-flowering tissues…….30

4.14. *In silico* identification and comparison of *cis*-acting elements present in the *COA* and *COB* homologs……..31

Objective 3: Identification and transgene mediated functional characterization of *BtFD1* and *BtFD2* gene homologs of *B. tulda* Roxb.

4.15. Identification of *FD1* and *FD2* gene homologs in the sequenced Poaceae and non-Poaceae genomes……..32

4.16 Phylogenetic analysis and comparison of *BtFD1* and *BtFD2* gene structure with their homologs present in the Poaceae and non-Poaceae plants…….35

4.17. Gene expression analyses of *BtFD* genes across five different seasons…35

4.18. *In silico* interaction study of *BtFD* gene homologs with the promoter of *AP1*……..36

4.19. Study of phenotypic characters of *35Sp::BtFD1* and *35Sp::BtFD2* transgenic plants……..……37

5. Discussion……..38-40

6. Literatures cited……...……41-44

7. Appendix 1

Plagiarism Certificate form Turnitin

Certificates of conferences attended

First pages of published manuscripts

1. Introduction

Bamboos are the gigantic member of grasses and grows in tropical as well as temperate regions (Trujillo et al., 2007). Approximately 1670 species of bamboo had been identified across the world (Soreng et al., 2017). Majority of the bamboos possess extensively long clonal growth phase, which is up to 120 years (Janzen, 1976; Tian,1987; Qin et al.,1989; Taylor et al., 1991). Another important character is monocarpy or semelparity where each and individual culm die after flowering. It is commonly believed as a bad omen since excessive seed settings in bamboo causes a sudden increase in rodent population which in turn causes great crop loss (John and Nadgauda, 2002). Therefore, in depth mechanistic understanding about the flowering in bamboo is very crucial for human and wild-life.

At a defined stage of the entire life history, plants transit from vegetative to flowering phase (Blackman, 2017). Therefore, plants have evolved a complex mechanism to ensure maximum reproductive success by inducing flowering at an appropriate time (Tremblay and Colasanti 2006). This transition of flowering is an irreversible process (Boss et al., 2004). During the transition phase, the shoot apical meristem (SAM) is transformed into either inflorescence meristem (IM) or floral meristem (FM), which result into severe changes in the growth and patterning of the main axis. The transition of flowering is two steps process i.e. induction and differentiation. In the induction step the plants commit for flowering. In the next step the induced plant evokes floral organ differentiation, i.e., production of flowers by promoting expression of the organ development genes. The decision of phase transition is not a random event. Studies over decades have affirmed that this is induced by the crosstalk between several developmental (phytohormones, plant age) as well as environmental (light, temperature) cues.

In the photoperiodic pathway the internal timekeeping mechanism of the plants is known as circadian clock. The circadian clock components are regulated by light signals perceived by the upstream photoreceptors such as *PHYTOCHROME*S (*PHY*) and *CRYPTOCHROME*S (*CRY*). The circadian clock is mainly composed of the genes namely, *TIMING OF CAB EXPRESSION 1* (*TOC1*) and *LATE ELONGATED HYPOCOTYL* (*LHY*) or *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), *GIGANTEA* (*GI*) and *ZEITLUPE* (*ZTL*; **Figure 1**). The central oscillation of the circadian clock is governed by *TOC1* and *LHY* (Alabadi et al., 2001). These two gene regulate the downstream gene *GIGANTEA* (*GI*; Alabadi et al., 2001). The *LHY* expression is upregulated in the morning which causes the repression of of the *TOC1* by binding to its promoter. On the

other hand, *TOC1* upregulates the expression of *LHY* and suppresses *GI* (Alabadi et al., 2001; Ito et al., 2008; **Figure 1**). In contrary, blue light stabilizes the interaction among ZTL and GI (Kim et al., 2007), which causes degradation of TOC1 at the post-translational level (Kim et al., 2007). Finally, the GI passes the circadian output to the master clock integrator *CONSTANS* (*CO*;

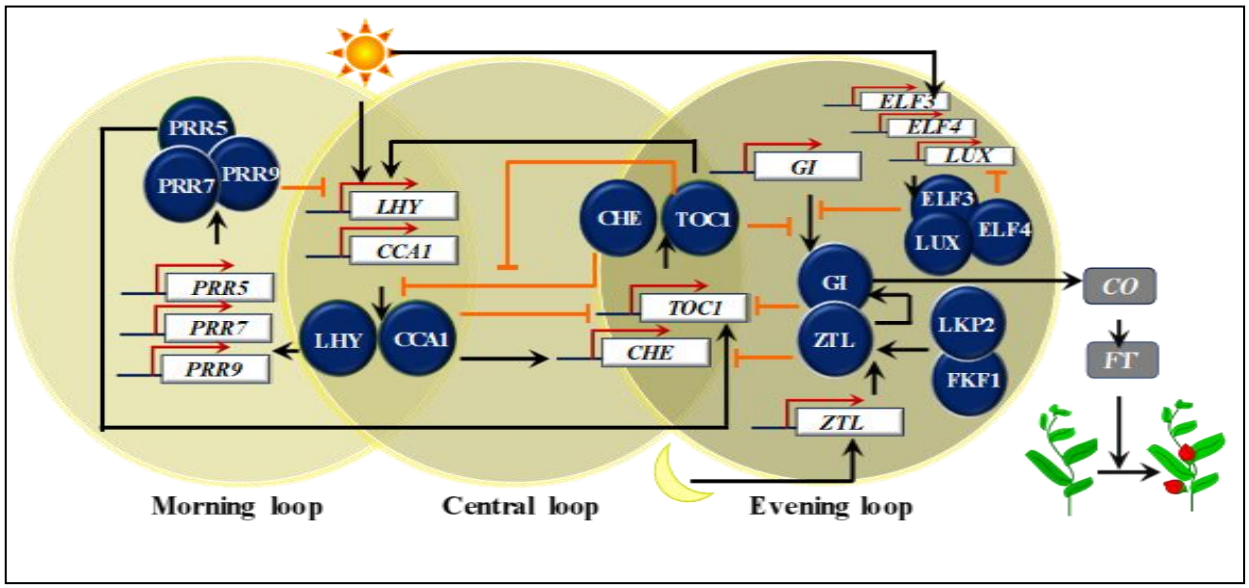

Figure 1. Schematic representation of transcriptional and translational feedback loop of the genes involved in photoperiodic pathway of *A. thaliana*. The morning loop is represented by the regulation between *LHY* and *PRR* gene family members. The central loop acts as the connecter between morning and evening loop. The evening loop is critical for activation of *CO*. Arrows indicate activation and T-bars indicate repression of gene activity.

Herrero et al., 2012). The *CO* is the circadian clock integrator gene, which accepts the photoperiodic signal coming from the clock genes and upregulate *FT* for the flower induction (Putterill et al., 1995). *CO* is the member of B-box family Zn finger protein (Griffith et al., 2003). Depending on the number of B-boxes *CONTANS LIKE* (*COL*) genes can be classified into four major groups. Other than the B-box region, the conserved CCT domain region, present at the carboxyl terminal end of the protein (Griffith et al., 2003) are involved in the interaction with HEME ACTIVATOR PROTEIN (HAP) transcriptional activators (Ben-Naim et al., 2006; Wenkel et al., 2006). This specific interaction is essential for the transcriptional activation of *FT* (Ben-Naim et al., 2006). In *Arabidopsis*, FT interact with FD at the shoot apical meristem region (Abe et al., 2005; Wigge et al., 2005). *FD* is a basic leucine zipper (bZIP) family transcription

factor which can interact with the ACGT motif located in the promoter of *AP1* to induce flowering (Li and Dubcovsky, 2008; Tsuji et al., 2013).

Genetic studies on flowering have predominantly been done on the model dicot plant *Arabidopsis thaliana* and the monocot plant *Oryza sativa* (Putterill et al., 1995). However, both these plants are annual and therefore further studies are required to understand these pathways and genes in the perennial plants, where flowering happens after many years of vegetative growth of the plant. Bamboo is one such extreme example of perennialism. The main aim of this study is to understand the regulation of the important circadian clock genes and the possible involvement of the integrator genes in controlling flowering time in the tropical bamboo, *B. tulda* Roxb.

2. Objectives (as specified in approved 1000 word):

Objective 1. Identification and molecular characterization of important circadian clock genes in bamboo

- *In silico* identification of important circadian clock genes in *Phyllostachys heterocycla*
- *In silico* identification of *cis*-acting elements obtained from promoters of the selected circadian clock genes of *Phyllostachys heterocycla* comparison to that of related monocot sequences
- PCR isolation, sequencing and detailed molecular characterization of four important circadian clock gene homologs (*BtLHY, BtTOC1, BtZTL, BtGI*) in *B. tulda* Roxb.
- Study of evolutionary relationship of the four *B. tulda* circadian clock genes, with that of related monocot species by phylogenetic analyses
- Study on the tissue specific expression patterns of the circadian clock genes by real time RT-qPCR analyses in *B. tulda*
- Study on the diurnal expression patterns of the circadian clock genes by real time RT-qPCR method in short day and long day conditions in *B. tulda*

Objective II. Identification, sequencing and detailed molecular characterization of important circadian clock integrator gene, *CONSTANS* in *B. tulda* Roxb.

- PCR isolation, sequencing and molecular characterization of two *CONSTANS* gene homologs (*BtCOA, BtCOB*) in *B. tulda* Roxb.
- Study of evolutionary relationship of the *BtCOA, BtCOB* in comparison to that of related monocot species

- Study on the tissue specific gene expression patte*rns* of the *BtCOA, BtCOB g*ene homologs by real time RT-q*PCR metho*d
- Study on the diurnal expression patterns *BtCOA, BtCOB g*ene homolog*s by real* time RT-qPCR method in short day and long day
- *In silico* study on the molecular interactions between individual BtCO and HAP-3 proteins

Objective III. Identification and transgene mediated functional characterization of *BtFD1* and *BtFD2* gene homologs of *B. tulda* Roxb.

- Sequence characterization and phylogenetic analyses of *BtFD1* and *BtFD2* among the Poaceae and non-poaceae members of the related monocot species
- Study on the seasonal changes of the expression patterns of *BtFD1, BtFD2* homologs by real time RT-qPCR method in *B. tulda*
- Gateway cloning and *in planta* transformation and detailed phenotyping analyses of the *BtFD1, BtFD2* overexpressed *A. thaliana* Col-0 plant
- Study on the *in silico* interaction of the *BtFD* gene homologs with the promoter of *AP1* gene

3. Methodology

3.1. Collection of tissues for isolation of nucleic acids

Different flowering and non-flowering tissues were collected from the three identified flowering populations located in Shyamnagar, W.B. (SHYM7, SHYM16, 22.38º N. 88.40º E) and Bandel (BNDL22, 22.93º N. 88.38º E) of *B. tulda*. Healthy, young leaf tissues were harvested for isolation of genomic DNA and RNA. To analyze expression of the targeted genes in different tissues, seven non-flowering tissues, i.e., leaf from flowering (PFL) and non-flowering branches (YLF) of a flowering culm, leaf from non-flowering culm (YLN), flag leaf (FL), culm sheath (CS), rhizome (RH), root (R) and internode (IN) and three flowering tissues, i.e., early (E), middle (M) and late (L) staged inflorescence buds were obtained (**Table 1**). The diurnal gene expression study was conducted in YLF and YLN, which were selected as representative tissues of floral inductive and non-inductive stages respectively in four different time points of a day (morning- 8 am, noon- 12 pm, afternoon- 4 pm and night- 8 pm) under short day (SD, 11 h light, sunrise at 6 am and sunset at 5 pm) and long day (LD, 14 h light, sunrise at 4:30 am and sunset at 6:30 pm) conditions. YLF and YLN were also collected in five different seasons- summer (April-June, 2017), monsoon (July-August, 2017) autumn (September-October,

Table 1. Summary of tissue samples collected for nucleic acids isolation for diverse experimental purposes

Purpose of the experiment	Timing of tissue collection	Plant tissues collected
Tissue specific expression analyses	Throughout the year	Leaf from flowering (PFL) and non-flowering branches (YLF) of a flowering culm, leaf from non-flowering culm (YLN), flag leaf (FL), culm sheath (CS), rhizome (RH), root (R) and intermodal (IN) region for non-flowering tissues. Early (E), middle (M) and late (L) staged inflorescence buds for flowering tissues
Diurnal expression analyses	Long day: June-July, Short day: December-January	Leaf from non-flowering branches of a flowering culm (YLF) and leaf from non-flowering culm (YLN)
Seasonal expression analyses	Summer: May-June, Monsoon: July-August, Autumn: September-October, Winter: November-January, Spring: February-April	Leaf from non-flowering branches of a flowering culm (YLF) and leaf from non-flowering culm (YLN)

2017), winter (November-January, 2017) and spring (February-March, 2018). Three biological replicates of each tissue stages were used for gene expression analysis. However, due to unavailability of the sufficient flowering tissues only one replicate was considered for the diurnal experiment under SD condition. The voucher specimens were deposited to the Botanical Survey of India (B.S.I), Shibpur (deposition nos. 56A, 56B, 57A, 57B, 58A. 58B, 59A, 59B, 59C on 05.06.2015).

3.2. Retrieving genomic and protein sequences from different online sequence resources

In order to design primers and conduct in depth sequence characterization of the genes isolated from *B. tulda* in comparison with related monocotyledonous plant species genomic and amino acid sequences were retrieved by BLASTP analyses in phytozome, plaza_monocot_v4 and NCBI by using the homologs of *O. sativa* as query (**Table 4, 5, 6**). The set criteria for BLASTP analysis were an E-value: $\leq e^{-10}$, identity: ≥40% and query coverage: ≥40%. When multiple hits were obtained, only the best BLASTP hit was considered (**Table 4, 5, 6**). However, if no homologs were identified, it is mentioned as no hit found (NHF, **Table 4, 5, 6**).

3.3. Isolation of genomic DNA

Genomic DNA isolation was carried out from the leaves by using the DNeasy Plant Mini Kit (Qiagen, Germany). Approximately 600mg tissues were taken and the sample was treated with RNaseA (100mg/ml, Thermo Scientific, USA). Ultimately, the genomic DNA was eluted by adding 50µl of DNase, RNase and protease free water and the quantity and quality of the isolated genomic DNA were observed in a BioSpectrometer (Eppendorf, Germany) and 1% low EEO agarose (SRL, India) gel electrophoresis.

3.4. Designing of primers for gene amplification and expression analysis

Homologous genomic DNA and/ cDNA sequences obtained from related monocot species were multiple aligned in MUSCLE program. The conserved regions were identified in order to design the gene specific degenerate primers to amplify the genes (**Table 2)**. The gene specific primers were designed from homologs isolated from *B. tulda* to carry out the gene expression analysis by the real time RT-qPCR. All the primers were designed in Primer3 web tool.

Table 2. Summary of primers designed for gene amplification and real time RT-qPCR.

Target genes	Oligo names	Designed from	Sequences (5'-3')	Length (bp)	Tm (^{o}C)	G/C (%)	Experimental purposes
LHY	LHY_DG_F	Conserved regions of related species	AGTTGGARAAGGAAGCTATCAAYAATG	27	60.52	37.04	PCR amplification of the target gene
	LHY_DG_R	Conserved regions of related species	TCRTCACTAGCCGGCACCC	19	61.20	63.16	
	LHY_5´_HD_F	Conserved regions of related species	cggtacccggggatcATGGAGATWAATTCTTCKGGTG	37	>75	56.75	PCR amplification of the 5´ end of the target gene
	LHY_5´_HD_R	*BtLHY*	cgactctagaggatcAGATATGGACTGTTAGGTTTTCTT	39	>75	41.02	
	LHY_3´_R	Conserved regions of related species	TCAGGTGGACGCTTCGCTCTC	21	64.41	61.90	PCR amplification of the 3´ end of the target gene
	LHY_QRT_F	*BtLHY*	CACAAGTGATTCATGGAAGGAAGTTTC	27	61.00	40.74	Gene expression analyses by real time RT-qPCR
	LHY_QRT_R	*BtLHY*	GGCGTTCTTGTTGAGGTCGA	20	60.60	55.00	
TOC1	TOC1_DG_F	Conserved regions of related species	ATGGTGGGCACCAGCGAG	18	62.42	66.67	PCR amplification of the target gene

	TOC1_DG_R	Conserved regions of related species	CTACTCTGGDGAAGAAACCATCTCTACCTC	30	63.55	46.67	
	TOC1_QRT_F	*BtTOC1*	AGGAACAAGGAACTGCGCCACATC	24	64.19	50.00	Gene expression analyses by real time RT-qPCR
	TOC1_QRT_R	*BtTOC1*	GTCTTCGCCGCCACACATG	19	64.84	68.42	
ZTL	ZTL_DG_F	Conserved regions of related species	TGCAATGYAGAGGVCCGTTTG	21	61.16	52.38	PCR amplification of the target gene
	ZTL_DG_R	Conserved regions of related species	TTCRGTTAGCATCCACTCTTCTCC	24	60.32	45.83	
	ZTL_3´_R	Conserved regions of related species	TCAAACTGAGCTGCTAGCCAGAG	23	62.49	52.17	PCR amplification of the 3´ end of the target gene
	ZTL_QRT_F	*BtZTL*	TGTCACCCAGGGATATAGCATC	22	59.09	50.00	Gene expression analyses by real time RT-qPCR
	ZTL_QRT_R	*BtZTL*	TGGTCAACTCTCTTGCTAGTCG	22	59.77	50.00	
GI	GI_5´_HD_F	Conserved regions of related species	cggtacccggggatcATGTCAGCTTCAAATGAGAAGTGG	39	>75	53.84	PCR amplification of the 5´ end of the target gene
	GI_5´_HD_R	Conserved regions of related species	cgactctagaggatcTCTAAGGAGTTCAACAAGCTG	36	>75	47.22	
	GI_3´_HD_F	Conserved regions of related species	cggtacccggggatcCTACTCCAAGTGCTACGCAGAGGTT	40	>75	60.00	PCR amplification of the 3´ end of the target gene
	GI_3´_HD_R	Conserved regions of related species	cgactctagaggatcTCAGCAAGKGAGWGGGCAGCC	36	>75	63.88	
	GI_QRT_F	*BtGI*	GAACCATATGCTCGCTTGTT	20	56.50	45.00	Gene expression analyses by real time RT-qPCR
	GI_QRT_R	*BtGI*	TCTAAGGAGTTCAACAAGCTG	21	55.55	42.86	
COA	COA_DG_F	Conserved among related species	TRYGAHGCRVAGGTGCAC	18	63.9	66.6	PCR amplification of the target gene
	COA_DG_R	Conserved among related species	TRGCGAAVCGDCCCTTGAYC	20	63.4	60	
	COA_5´_F	Conserved among related species	ATGAATTATAATTTCARCAGCHMCGT	26	57.7	30.7	PCR amplification of the 5´ end of the

	COA_5´_R	*BtCOA*	ACACTCGCTCCTTTC CTTCTCTG	23	59.6	47.8	target gene
	COA_3´_F	*BtCOA*	TATGATTTGCACATT TGAGTGACCATTT	28	60.7	32.1	PCR amplification of the 3´ end of the target gene
	COA_3´_R	Conserved among related species	TCAGAACCATGGAA CGGTRCYRTA	24	64.1	50.0	
	BtCOA_Q RT_F	*BtCOA*	CTGCTGCTCAACAAA GATTCT	21	56.8	42.8	Gene expression analyses by real time RT-qPCR
	BtCOA_Q RT_R	*BtCOA*	CTCTTGGTTGTTGTC GATGTG	21	57.4	47.6	
COB	COB_DG_ F	Conserved among related species	CGAGCCGGAGCTCCS YGGCC	20	68.2	75	PCR amplification of the target gene
	COB_DG_ R	Conserved among related species	GCCGGGCACGACGA RCTCGTG	21	68.5	71.4	
	COB_3´_F	*BtCOB*	CCGTTCGGCGACGC ACTG	18	60.6	66.6	PCR amplification of the 3´ end of the target gene
	COB_3´_R	Conserved among related species	CTAGAACGTCGGCA CGAC	18	58.0	61.1	
	BtCOB_Q RT_F	*BtCOB*	GCTGCGACTCGTGCC GTTC	19	64.4	68.4	Gene expression analyses by real time RT-qPCR
	BtCOB_Q RT_R	*BtCOB*	CAAGAGGAGTTGAC ACCGGGC	21	63.2	61.9	
FD1	BtFD1_QR T_F	*BtFD1*	GCAGGAAAACGAGA AGCTCCGTGTCA	26	67.30	53.85	Gene expression analyses by real time RT-qPCR
	BtFD1_QR T_R		TCAGAATGGCGCCG AGAGC	19	62.14	63.16	
	BtFD1_G W_F	*BtFD1*	ggggacaagtttgtacaaaaaag caggcttaATGGCCATGG AGGACGAC	49	>75	49	PCR amplification for Gateway cloning
	BtFD1_G W_R		ggggaccactttgtacaagaaag ctgggTCAGAATGGCG CCGAGAG	46	>75	56.5	
	BtFD1_F	Conserved among related species	ATGGCCATGGAGGA CGACG	19	62.14	63.16	Molecular confirmation of transgenic plants
FD2	BtFD2_QR T_F	*BtFD2*	ATGGCGGCCAACTA CCACCACT	21	67.01	59.09	Gene expression analyses by real time RT-qPCR

	BtFD2_QRT_R		CGGTTCCTCATCATCCTGATGG	22	66.02	54.55	
	BtFD2_GW_F	*BtFD2*	ggggacaagtttgtacaaaaaagcaggcttaATGGCGGCCAACTACCAC	49	>75	49	PCR amplification for Gateway cloning
	BtFD2_GW_R		ggggaccactttgtacaagaaagctgggTCAGAATTGCGTCGATGAGG	48	>75	52.1	
	BtFD2_R	Conserved among related species	TCAGAATTGCGTYGATGAGGTTC	23	65.76	47.83	Molecular confirmation of transgenic plants

3.5. Amplification of the targeted flowering genes by PCR

PCR amplifications were carried out by using Phusion Taq DNA polymerase (Thermo Scientific, USA) in S1000 PCR detection system (Bio Rad, USA). The final concentrations of the components of PCR master mix were: 1.5-2.5mM $MgCl_2$, 0.2mM of dNTPs, 0.25µM forward, and 0.25µM reverse primer, 1.5ng/µL of genomic DNA along with 0.02U/µL of phusion DNA polymerase (Thermo Scientific, USA) or 1.25 U/50µL of extaq DNA polymerase (Takara, USA). The PCR reactions were performed using the following reaction cycles: denaturation at 98 °C for 30 seconds, 25-35 cycles [98 °C for 10 seconds, 55-65 °C for 30 seconds, 72 °C for 30-60 seconds per kb of PCR product] and the final extension at 72 °C for 5 minutes.

3.6. Calcium competent cells (*E. coli Dh5α*) preparation

In order to perform molecular cloning *E. coli Dh5α* calcium competent cells were prepared. Firstly, the starter culture of overnight grown *E. coli Dh5α* cells were inoculated in 250 ml of liquid LB medium and incubated at 37 °C. When the bacterial cell density (OD_{600}) reached at 0.35 to 0.40, cell pellets were collected by successive centrifugation steps and treated with chilled 100mM $MgCl_2$, 100mM $CaCl_2$ and 85mM $CaCl_2$ in 15% glycerol solution. Finally, the competent cells were suspended in 85mM $CaCl_2$ in 15% glycerol and snap frozen in liquid nitrogen.

3.7. Gel elution of the amplified PCR products having desired molecular weight

GeneJET gel elution kit (Thermo Scientific, USA) was used for elution of the PCR amplified bands of desired molecular weight. Approximately 100 to 150 µL of PCR products were used for

this purpose. The desired bands were excised and suspended immediately into equal volume of binding buffer and kept at 55 - 60 °C. After passing the solution through DNA binding columns, the PCR products were eluted with 20 µL nuclease free water. Molecular weight of the eluted PCR product was confirmed by running the product on 1.5% agarose gel (SRL, India) gel.

3.8. Cloning of the gel eluted PCR products and transformation

PCR products having 3' A-tail were cloned by using pGEM®-T Easy vector systems (Promega, USA), while for the blunt ended products CloneJET PCR Cloning Kit (Thermo Fisher Scientific, USA) was used. For cloning longer PCR products of length >2000 bp, infusion HD cloning kit was used (Takara, USA). Approximately, 25ng of vector DNA was ligated with the eluted PCR products in 1:3 ratios when pGEMT and CloneJET PCR Cloning Kit was used. However, approximately 100 ng of PCR product was used in case of infusion HD cloning kit as per the recommendation of the manufacturer. Finally, the transformed cells were plated onto solid LB media containing 1.6% agar and ampicillin (100µg/ml) as selectable marker. In addition, for TA and In Fusion HD cloning, agar plates were supplemented with X gal (20mg/ml, dissolved in formamide) and IPTG (0.1M). The plates were incubated at 37 ^{o}C overnight to promote bacterial growth.

3.9. Colony PCR

Randomly about 10 - 20 single colonies were subjected to colony PCR. The PCR reaction conditions used were as per section 3.5.

3.10. Isolation of plasmids and Sanger's sequencing

Bacterial colonies containing the right insert were inoculated to 10 ml of liquid LB medium containing Ampicillin (100µg/ml) as the selectable marker. The GeneJET Miniprep plasmid isolation kit (Thermo Scientific, USA) was used for plasmid isolation. At first, bacterial pellet was harvested by centrifugation using 5ml bacterial solution. Resuspended bacterial cells were subjected to alkali lysis and successive neutralization using appropriate buffers provided with the kit. Finally, plasmids were eluted in 30µl of nuclease free water and sent for standard Sanger's sequencing (AgriGenome, Kochi). The trimmed nucleotide sequences were then used to obtain the contig using CAP3 program and the final sequences were deposited to NCBI.

Table 3. Summary of the databases/webtools used for various *in silico* analyses.

Purpose of the experiment	Database/Web tool	Weblink	Version
Retrieving genomic, amino acid sequences of *O. sativa*	Rice genome annotation project	http://rice.plantbiology.msu.edu/	Release7
BLASTP analysis	Phytozome12	https://phytozome.jgi.doe.gov/pz/portal.html	v12.1.5
BLASTP analysis	Monocot plaza	https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_monocots/	v4.0
BLASTP analysis	National Center for Biotechnology Information	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins	-
Multiple sequence alignment	MUSCLE	www.ebi.ac.uk/Tools/msa/muscle/	August 2004
Primer designing	Primer3 web tool	http://primer3.wi.mit.edu/	Release 0.4.0
Gene sequence assembly	CAP3	http://doua.prabi.fr/ software/cap3	Jan 2014
Prediction of gene models	Gene Structure Display Server	http://gsds.cbi.pku.edu.cn/index.php	v2.0
Multiple sequence alignment	Clustal omega	https://www.ebi.ac.uk/Tools/msa/clustalo/	-
In silico amino acid characterization	ExPASy (ProtParam)	https://www.expasy.org/	-
In silico identification of *cis*- acting elements	PlantCARE database	http://bioinformatics.psb.ugent.be/webtools/plantcare/html/	-
Construction of dendrograms	DendroUPGMA	http://genomes.urv.cat/UPGMA/index.php?entrada=Example1	November 2016

3.11. Prediction of gene models and *in silico* characterization of predicted protein sequences

The gene models were predicted by using the full length genomic and coding sequences (CDS) in the web tool Gene Structure Display Server (**Table 3).** The predicted protein sequences were multiple aligned in clustal omega tool and the signature domains and the biologically important residues were marked. The predicted protein sequences were used to calculate the molecular weight, isoelectric point, instability index, aliphatic index and grand average of hydropathicity (GRAVY) using tools available in the ExPASy database.

3.12. Phylogenetic analyses

The full length protein sequences sequences were first multiple aligned using MUSCLE tool and then the unrooted, Neighbour Joining (NJ) tree was constructed in MEGA7 program by using the default parameters (Kumar et al., 2016). The bootstrap value used were 1,000.

3.13. Extraction of total RNA

Isolation total RNA from diverse tissue stages were carried out by following the standardized protocol combining TRIzol and RNeasy plant minikit (Qiagen, Germany, Das et al., 2010; 2016). To avoid any genomic DNA contamination, the lysate was subjected to DNase I enzyme treatment (Thermo Scientific, USA). Finally, quantity and quality of the isolated total RNA was evaluated in BioSpectrometer (Eppendorf, Germany) and agarose-formamide gel electrophoresis.

3.14. Preparation of cDNA libraries

The cDNA libraries were prepared with the help of Verso cDNA synthesis kit (Thermo Scientific, USA) by following the protocol provided in the user manual. Approximately, 1µg of total RNA was used per library in order to get the maximum yield of cDNAs. The oligo-dT and random-hexamer primers were used in 3:1 ratio. PCR cycling steps used were as follows: 25 ^{o}C for 10 min, 42 ^{o}C for 1h followed by a step at 70 ^{o}C for 15 min.

3.15. Real time RT-qPCR analyses

The real time RT-qPCR analyses were performed in CFX connect real-time PCR detection system (Bio Rad, USA) by using SsoAdvanced™ Universal SYBR® Green Supermix kit (BioRad, USA). For each set of reactions, 10µl reaction mixture was prepared. The final concentration of the components of reaction mixture were: supermix buffer 1x, each of the primers 0.25µM. In addition, 2µL of template cDNA (1/20th diluted from stock) was added to each reaction. The PCR amplification cycles used were: 30s at 95°C, 40 cycles [10 s at 94°C and 40 s at 55 °C]. A dissociation curve analysis was done to confirm the absence of any primer dimers. The relative fold change was calculated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The reference gene *eIF4α* was used for the data normalization.

3.16. *In silico* prediction of *cis*-elements and construction of dendrograms

In order to identify the important *cis*-acting elements possibly responsible for gene expression divergence, the PlantCARE database were searched by using 1500 bp upstream region from the

translation initiating site (ATG). Data obtained from the Plant Care database were analyzed and the total numbers were recorded for each cis-elements and was used as the raw data sheet in DendroUPGMA program to generate the dendrogram (**Table 3)**.

3.17. Gateway cloning and *Agrobacterium* mediated transformation of two *BtFD* gene homologs

The genomic DNA was used to amplify *BtFD1* and *BtFD2* genes by using by using their respective gateway primers along with Phusion taq DNA polymerase (Thermo sciectific, USA) and gel eluted. In the BP reaction, approximately 100ng of the PCR elutes were recombined with approximately 100ng of pDONR221 donor vector using BP Clonase™ enzyme mix and the BP Clonase™ reaction buffer (Invitrogen, USA). Approximately 5ul of the ligation cocktail was used for transforming Ca^{2+}-competent *E. coli* (*DH5α*) cells by heat shock method. The transformed cells were grown over night at 37°C on solid LB agar having Kanamycin as the selectable marker (50µg/ ml; Duchefa, Netherlands). Plasmids were isolated from the putatively transformed colonies and were confirmed by restriction digestion and Sanger's sequencing. The restriction enzymes used for confirmation of *BtFD1* were *PvuII, XhoI* and for *BtFD2* were *HindIII, NheI* (Thermo Fisher, USA).

Plasmids having the right inserts were subjected to LR reaction using the vector pAlligator2 having the 35S CaMV promoter. The ligation mixture was then transformed into Ca-competent *E. coli* (*DH5α*) cells and grown over night at 37°C in solid LB plate having Spectinomycin (100µg/ ml; Duchefa, Netherlands). Putatively transformed colonies were grown overnight at 37C in liquid LB medium. Insertion of right DNA sequences were confirmed by performing restriction digestions of plasmids. The restriction enzymes used for confirmation of *BtFD1* were *SacI, NcoI* and for *BtFD2* were *HindIII, PstI* (Thermo Fisher, USA).

Plasmids with the correct inserts were then transformed into *Agrobacterium* cells (pGV3101/pMP90RK) by electroporation using a Bio-Rad Gene Pulser™. Reaction settings used were capacitance: 250 µF, pulse controller: 400 Ω and electric tension: 1.25 V. Transformed *Agrobaterium* cells were plated on solid LB medium containing 100 µg/ml Rifampicin (Duchefa, Netherlands), 25 µg/ml Gentamycin and 100 µg/ ml Spectinomycin.

3.18. *In planta* transformation of *BtFD1* and *BtFD2* genes under 35S promoter in *Arabidopsis thaliana* Col-0

A. thaliana ecotype Col-0 plants were transformed with p35S*:BtFD1* and p35S*:BtFD2* constructs separately. In order to do so, positively transformed *Agrobacterium* cells were grown overnight at 28^{o}C in 2 ml of liquid LB medium having 25 µg/ml Gentamycin, 100 µg/ml Rifampicin and 100 µg/ ml Spectinomycin. Next day, 1 ml of this pre-culture was added to 300 ml of liquid LB medium supplemented with same antibiotic compositions and were grown overnight. Next day, bacterial pellets were collected by centrifugation at 5000 rpm for 10 mins. and dissolved in 0.05% silwet solutions containing 5% sucrose. When the OD_{600} of the solution was around ≥0.8, approximately 4 weeks old, healthy *A. thaliana* plants having many unopened flower buds were dipped for 30 secs into this solution.

3.19. Selection and molecular confirmation of putatively transformed plants

Harvested seeds from the putatively transformed *A. thaliana* ecotype Col-0 plants (T_0) were observed in a UV fluorescence microscopy (Leica, Carl Zeiss) and seeds emitting strong green fluorescence were grown in soil for subsequent phenotypic observations. Molecular confirmation of the putatively transformed plants was performed by PCR amplification of *BtFD1* and *BtFD2* gene using gene specific primers.

3.20. Phenotypic analyses of the transgenic plants

All the phenotyping experiments were carried out by growing the plants in plant growth chamber in long day (16h light/ 8h dark, 22-24^{o}C) and short day (10h light/14h dark) conditions. Leaf numbers and flowering time of the transgenic plants were recorded and compared to that of wild type plants.

3.21. *In silico* interaction of *BtFD1* and *BtFD2* proteins with the promoter of *AP1* gene

Preliminary sequence analyses revealed that bZIP domains of both *BtFD1* and *BtFD2* belong to CREB subfamily that specifically bind to TGACGT DNA sequence present in the promoter region of AP1 gene. Since crystal structures of basic leucine zipper (bZIP) domains of *BtFD1* and *BtFD2* were unavailable, their respective sequences were subjected to homology modelling by SwissModel. Both of them demonstrated significant homology with the template, 1DH3, which is the crystal structure of cAMP responsive element-binding protein (CREB) in complex with CRE DNA. Ramachandran analysis revealed 98.96% residues were in favourable region for *BtFD1* and 95% for *BtFD2* and both modelled in dimeric state. The two bZIP models were then subjected to energy minimization using Chimera (https://www.cgl.ucsf.edu/chimera/). The

energy minimized structures were then docked with 21 bp CRE DNA sequence (obtained from 1DH3 crystal structure) using NPdock.

4. Results

Objective 1: Identification and molecular characterization of important circadian clock genes in bamboo

4.1. *In silico* identification of four important circadian clock genes from all the sequenced genomes of monocotyledonous plants

At least one gene homolog for each of *OsLHY*, *OsTOC1*, *OsZTL* and *OsGI* could be identified from majority of the Poaceae as well as non-Poaceae monocotyledonous species (**Table 4**). Although these four circadian homologs could be identified in the perennial temperate bamboo genome *P. heterocycla*, their sequence homology to that of *O. sativa* homologs varied. For instance, amino acids sequence similarity of the *P. heterocycla LHY* was 80.52%, *P. heterocycla TOC1* was 85.76%, *P. heterocycla ZTL* was 95.14% and *P. heterocycla GI* was 89.52%. These results suggested that the core clock members were conserved in both Poaceae and non-Poaceae plants.

4.2. Identification and molecular characterization of *LHY, TOC1, ZTL and GI* homologs in *B. tulda*

In this study four circadian genes (*LHY, TOC1, ZTL* and *GI*) have been isolated from *B. tulda* by designing gene specific degenerate primers and the sequences were deposited to NCBI database (*BtLHY*- MF983713, *BtTOC1*- KY249524, *BtZTL*- MF983715 and *BtGI*- MN167114). Their close homologs were also identified by BLASTP analysis performed in the non-redundant database of NCBI. For instance, highest homology for *BtTOC1* (85%), *BtLHY* (80.06%) and *BtGI* (91.73%) were obtained against the annual monocot species *O. sativa* (BAD38854.1), *O. brachyantha* (XP_006659145.1) and *O. sativa* (XP_015649578.1), respectively while *BtZTL* (93%) demonstrated highest sequence homology with the perennial *Thyridolepis multiculmis* (AML79118.1).

Table 4. *In silico* identification of circadian clock genes identified from monocotyledonous plant species. BLASTP analyses were performed using *O. sativa* amino acid sequences as quarries. Criteria used for BLASTP analyses and subsequent selection of homologs were: identities (≥40%), E values (e^{-10}) and and query coverage (≥40%). When multiple BLASTP hits were obtained, only top BLASTP hits were considered for further analyses.

Plant species	Flower-ing habit	Best BLAST hit			
		***LHY* homologs**	***TOC1* homologs**	***ZTL* homologs**	***GI* homologs**
O. brachyantha	Annual	OB08G13700	OB02G31620	OB06G33410	OB01G15220
Z. mays	Annual	ZM00001D049543	ZM00001D051114	ZM00001D014594	GRMZM2G107101T03
H. vulgare	Annual	HVU0045G1380	HVU0042G2124	HVU0045G1457	HVU0038G2147
S. bicolour	Annual	SOBIC007G047400	SOBIC004G216700	SOBIC004G042200	SOBIC003G0409003
T. aestivum	Annual	TAE56630G002	TAE14923G003	TAE06866G003	TAE09269G003
A. tauschii	Annual	XP0201954581	XP0201726651	XP0201966991	XP0201920411
B. distachyon	Annual	BRADI3G16515	BRADI3G48880	BRADI1G33610	BRADI2G052261
B. stacei	Annual	BRAST03G0181001	BRAST04G1362001	BRAST07G1973001	BRAST01G3600003
S. itallica	Annual	SEITA6G055700	SEITA1G236100	SEITA4G249100	SEITA5G1295001
S. viridis	Annual	SEVIR6G0531001	SEVIR1G2410001	SEVIR4G2616001	SEVIR5G1270001
P. hallii	Perennial	PAHALF010411	PAHALA025621	PAHALD004201	PAHALE042241
P. virgatum	Perennial	PAVIRFA019051	PAVIRAA013261	PAVIRDB002951	PAVIRJ132491
O. thomaeum	Perennial	OROPETIUM2015010523149	OROPETIUM2015010502777	OROPETIUM2015010514362	OROPETIUM2015010507316A
Z. japonica	Perennial	ZJNSC000711G025201	ZJNSC000171G004301	ZJNSC000941G009301	ZJNSC000111G064401
D. oligosanthes	Perennial	OEL370641	NHF	OEL289051	NHF
P. heterocycla	Perennial	PH01001283G0510	PH01003618G0130	PH01007024G0030	PH01002142G0290
Z. marina	Perennial	NHF	ZOSMA4G02070	ZOSMA25G00320	ZOSMA65G003901
S. polyrhiza	Perennial	SPIPO1G0078300	SPIPO0G0012200	SPIPO12G0055500	SPIPO14G0003100
M. acuminata	Annual	MAC11G0949	MAC08G0062	MAC03G0682	GSMUAACHR10T27330001
A. comosus	Perennial	ACO013228	ACO021178	ACO025121	ACO0143471
P. equestris	Perennial	PEQU07911	PEQU03015	PEQU31980	PEQU08319
E. guineensis	Perennial	EGU0206G1203	EGU0967G0013	EGU0881G0501	EGU0935G1751
P. dactylifera	Perennial	XP0088113971	XP0088057281	NHF	XP0087934641

4.3. Phylogenetic relationships of *LHY, TOC1, ZTL and GI* homologs of *B. tulda* with homologs identified from related monocotyledonous species

In order to understand the phylogenetic relationship of the four circadian clock genes (*BtLHY*, *BtTOC1*, *BtZTL* and *BtGI*) identified from *B. tulda* with that of related monocotyledonous species, NJ trees were constructed based on the protein sequences (**Figure 2, 3, 4, 5**). All the four

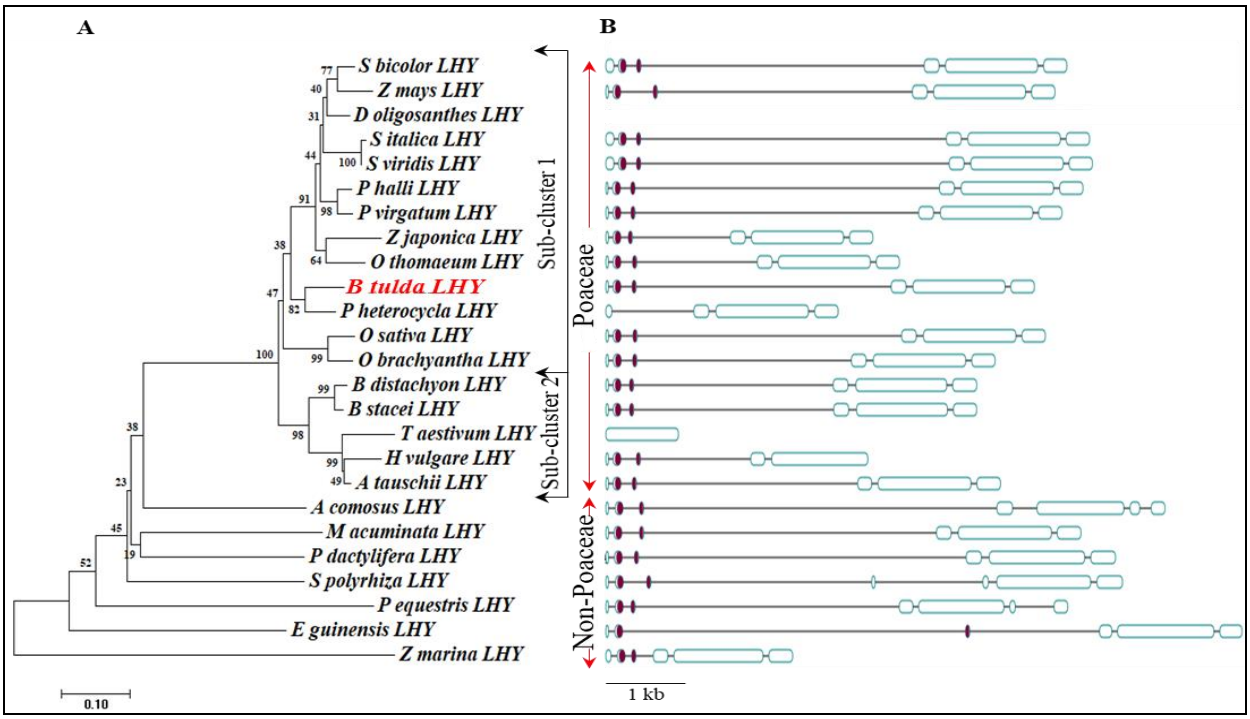

Figure 2. Phylogenetic analysis and gene model prediction of *LHY* genes identified in monocots. **A.** The Neighbour Joining (NJ) tree of the *LHY* homologs was constructed with the full-length protein sequences by Mega 7.0 in default parameters using 1000 bootstrap value. **B.** Predicted gene model of the identified LHY homologs. The blue cylinders indicate the exon. Introns are presented as black lines. Myb domains are marked in solid purple circles.

trees were grouped into two major clusters. The cluster 1 was comprised of the all Poaceae species ranging from annual crops to grasses along with two bamboo species (*B. tulda* and *P. heterocycla*), while cluster 2 contained homologs from all the non-Poaceae members such as *A. comosus, M. acuminata*, *P. dactylifera*, *S. polyrhiza*, *P. equestris*, *E. guineensis* and *Z. marina* (**Figure 2, 3, 4, 5**). In addition, the Poaceae specific cluster was again subdivided into two sub clusters in cluster 1 of *LHY, TOC1* and *GI* genes. The sub cluster 1 was composed of the annual (*O. sativa, Z. mays, S. bicolour, S. itallica, S. viridis*) as well as perennial Poaceae species (*O. thomaeum, Z. japonica, D. oligosanthes, P. hallii, P. virgatum*)**,** while sub-cluster 2 was only comprised of the annual plants such as *H. vulgare, T. aestivum, A. tauschii, B. distachyon* and *B. stacei*. Interestingly in case of the three homologs i.e., *LHY, TOC1* and *GI*, the identified *B. tulda* homologs were always grouped with another bamboo species i.e. *P. heterocycla* (**Figure 2, 3, 5**). In contrast, for *ZTL* the cluster 1 (Poaceae specific cluster) was sub-divided into two clusters and both these sub clusters hosted annual as well as perennial species (**Figure 4**). Interestingly it was

also observed that *BtZTL* remained in the subcluster 2 along with the *ZTL* homologs of *S. bicolor* and *O. thomeium*, which was quite distant from *P. heterocycla*. Therefore, it is evident from these observations that diversification of the amino acid sequences of *ZTL* homologs were more prominent than *LHY*, *TOC1* and *GI*.

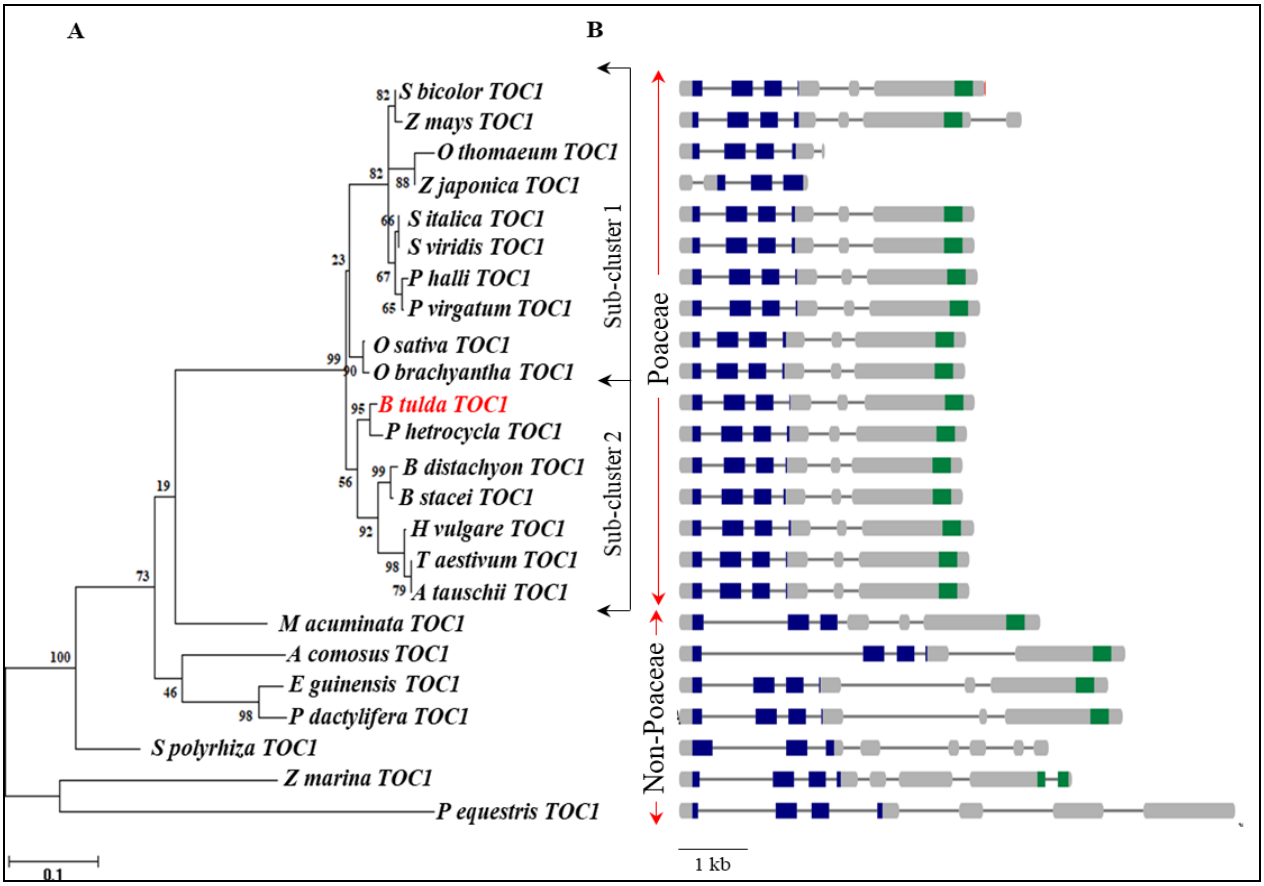

Figure 3. Phylogenetic analysis and gene model prediction of *TOC1* genes identified in monocotyledonous plants. A. Neighbour Joining (NJ) tree of the *TOC1* homologs was constructed with the full-length protein sequences by Mega 7.0 in default parameters using 1000 bootstrap value **B.** Predicted gene model of the identified *TOC1* homologs. The grey cylinders indicate the exon. Introns are presented as black lines. Receiver domains are marked in solid blue and CCT domains are in solid green rectangles.

4.4. Comparative analyses of gene structures and predicted amino acid sequences of *BtLHY, BtTOC1, BtZTL* and *BtGI* homologs with that of related monocotyledonous species

The detailed gene structure prediction revealed that the *BtLHY* gene is comprised of six exons and five introns which was quite comparable with the other identified *LHY* homologs (**Figure 2**). However, the exon numbers varied in most of the non-Poaceae species such as *A. comosus*, *P.*

equestris, *S. plyrhiza*, *Z. marina*. In addition, detailed sequence characterization revealed that *BtLHY* possessed a N terminal Myb domain, which was also found conserved except the temperate bamboo *P. heterocycla* (**Figure 2**). *In silico* characterization of the predicted protein sequences suggested that the predicted *BtLHY* protein was composed of 719 amino acids residues and the molecular weight was of 78598.22 kDa, which were comparable to the other *LHY* homologs identified from the Poaceae as well as non-Poaceae members.

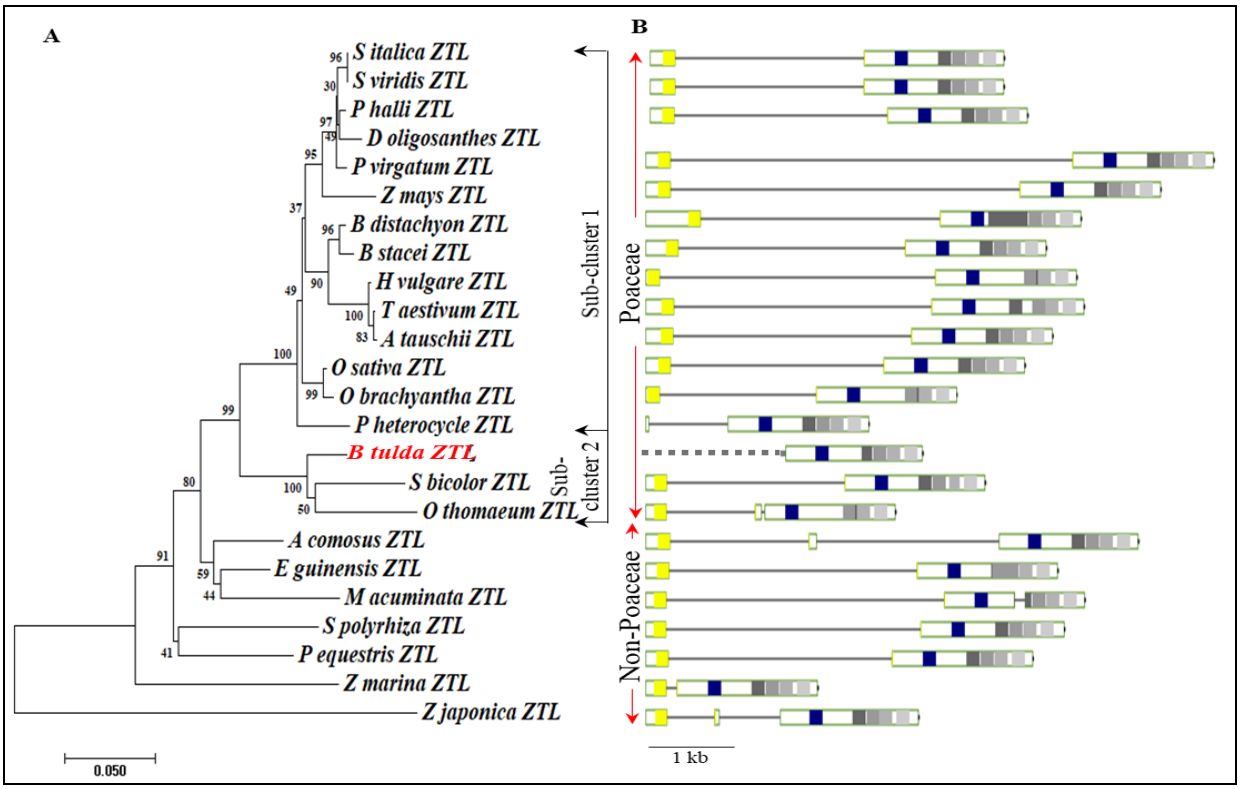

Figure 4. Phylogenetic analysis and gene model prediction of *ZTL* genes identified in monocots. A. Neighbour Joining (NJ) tree of the *ZTL* homologs was constructed with the full-length protein sequences by Mega 7.0 in default parameters using 1000 bootstrap value. **B.** Predicted gene model of the identified *ZTL* homologs. The green rectangles indicate the exon. Introns are presented as black lines. LOV, F-box and Kelch repeats are marked in solid yellow, blue and grey rectangles respectively.

Characterization of the gene sequences suggested that majority of the *TOC1* homologs identified from Poaceae species including that of *B. tulda* were comprised of six exons **(Figure. 3)**. *TOC1* is a member of *PRR* gene family. Analyses of the predicted protein sequences discovered that all the identified homologs including the *BtTOC1* possessed a N-terminal response receiver domain (RR; **Figure. 3**). Apart from RR domain the homologs also possessed a conserved CCT domain at their C-terminal end (**Figure. 3**). However, the CCT domain could not be found in the two

Poaceae (*O. thomaeum* and *Z. japonica*) and two non-Poaceae members (*S. polyrhiza* and *P. equestri*). Predicted protein sequence length and molecular weight of BtTOC1 were 522 and

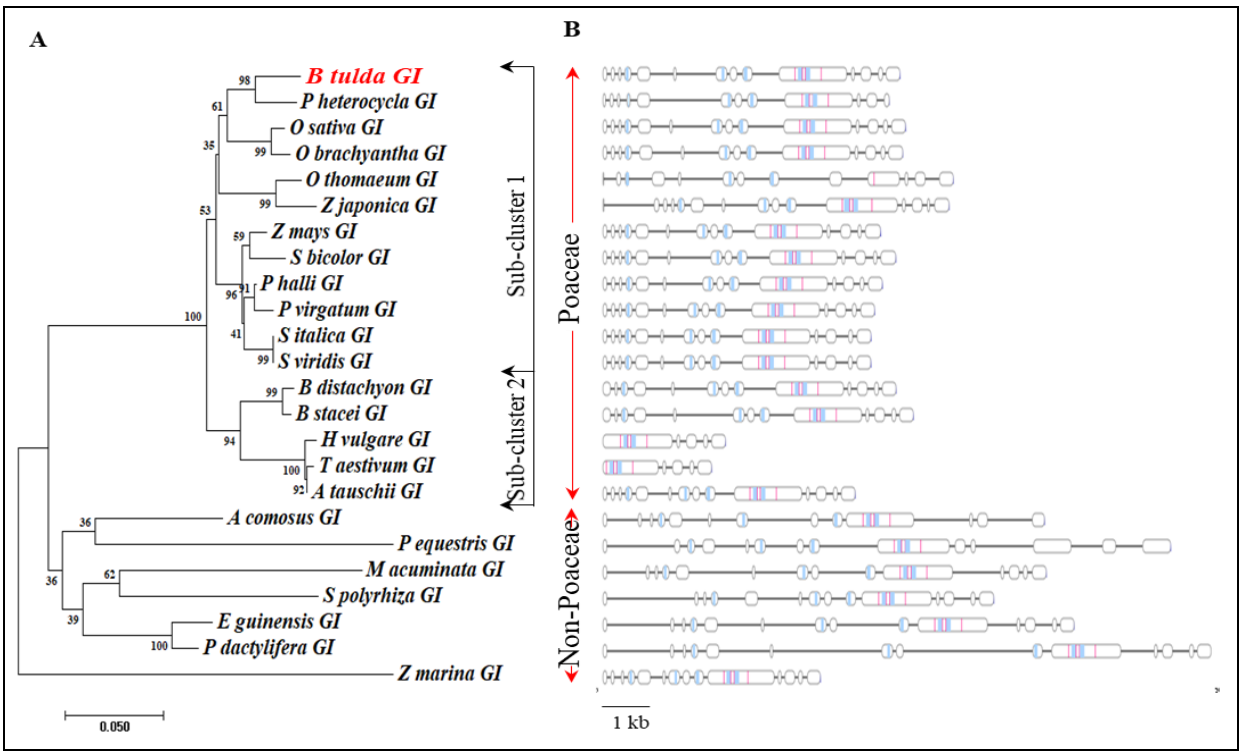

Figure 5. Phylogenetic analysis and gene model prediction of *GI* genes identified in monocots. A. Neighbour Joining (NJ) tree of the *GI* homologs was constructed with the full-length protein sequences by Mega 7.0 in default parameters using 1000 bootstrap value. **B.** Predicted gene model of the identified *GI* homologs. The grey cylinders indicate the exon. Introns are presented as black lines. The NLS and transmembrane domains are marked in solid pink and blue rectangles respectively.

58008.73kDa, respectively. Comparative analyses of the *ZTL* gene sequences identified in Poaceae and non-Poaceae species suggested that the homologs contain two exons and a single intron except in *O. thomaeum*, *A comosus* and *Z japonica* (**Figure 4**). *ZTL* is a F-box family gene and all the homologs including *BtZTL* contain a conserved F-box domain (**Figure 4**). Apart from the F-box domain, a conserved N- terminal light-oxygen-vol domain (LOV) and four kelch (kelch-1, 2, 3, 4) repeats were also observed which was also conserved among other identified *ZTL* homologs. *In silico* comparison of the amino acid sequences revealed that the molecular weight of these F-box proteins varied from 58096.42 to 80226.80 kDa and the lengths were 529 to 744.

Predicted gene model of *BtGI* revealed that the gene consisted of fourteen exons which was also the case in the reference species *O. sativa* (**Figure 5**). In contrast, number of exons were less in the temperate bamboo, *P. heterocycla*. Analyses of the predicted protein sequences demonstrated that the *BtGI* was composed of four nuclear localization domains (NLS1, 2, 3, 4) and five transmembrane domains (**Figure 5**). *In silico* study revealed that the molecular weight of BtGI was 126784.25 kDa.

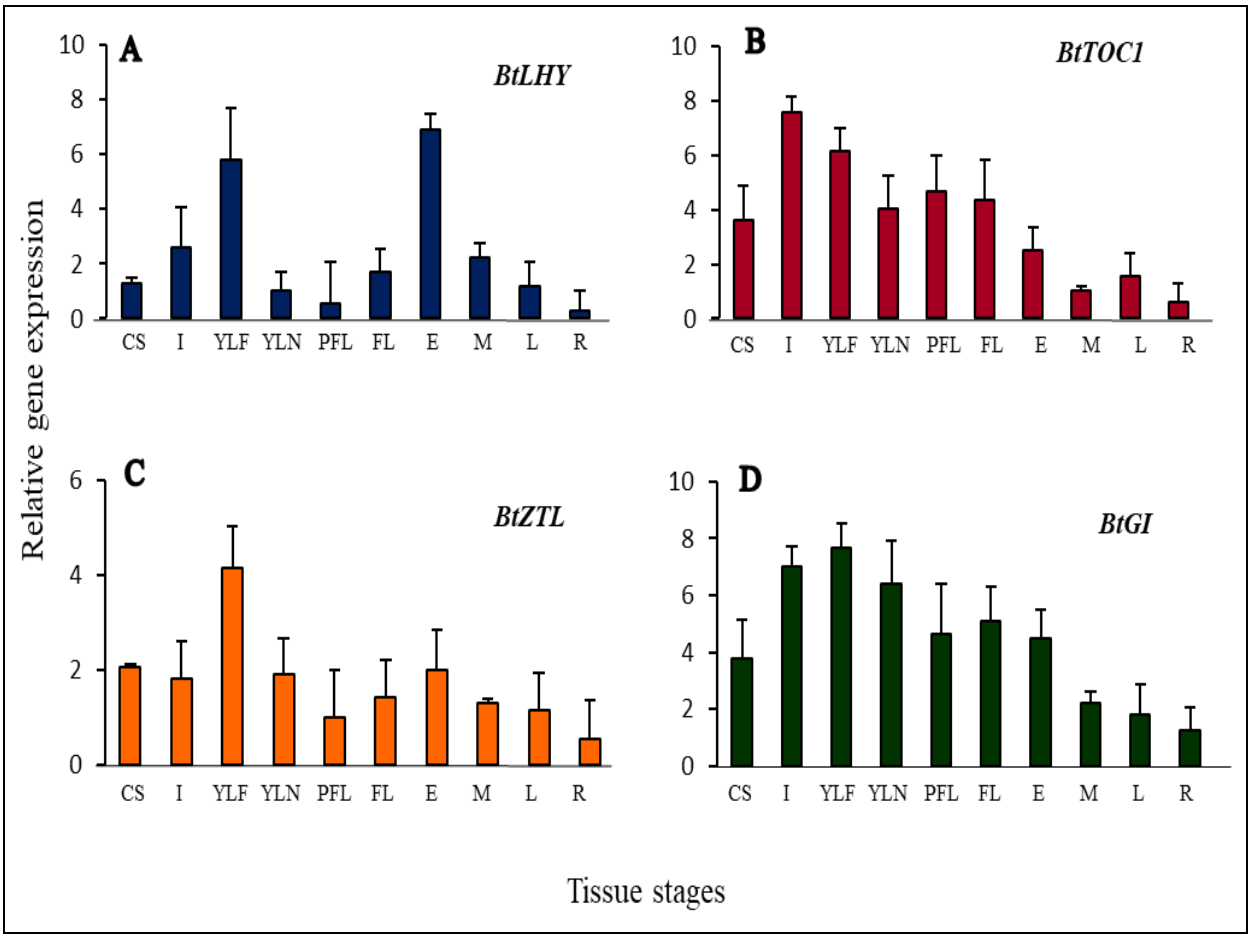

Figure 6. Gene expression analysis of circadian clock genes in *B. tulda*. The *eIF4α* was used as reference gene and rhizome as the calibrator. The $2^{-\Delta\Delta CT}$ method was used to measure the relative fold of the targeted genes. The fold change value of the three biological replicates ±SE were represented in each bar. Abbreviations used: CS: culm sheath, YLF- young leaf from flowering culm, I: internode, PFL: possible flag leaf, FL: flag leaf, E: early staged inflorescence bud, YLN- young leaf from non-flowering culm, M: middle staged inflorescence bud, L: late staged inflorescence bud, R: root.

4.5. Tissue specific gene expression analyses of *BtLHY, BtTOC1, BtZTL* and *BtGI* in *B. tulda*

Transcriptional accumulation of all of these four homologs were predominantly observed in the leaf tissues (**Figure 6**). Interestingly, differences in expression between the leaves collected from flowering (YLF) and non-flowering (YLN) plants were quite distinct in case of *BtLHY* and *BtZTL*, which was not the case for *BtTOC1* and *BtGI*. However, all these four genes revealed higher level of expression in YLF than YLN suggesting their roles in flowering. It was also observed that the highest level of expression of *BtLHY* and *BtTOC1* gene was in the early staged inflorescence buds (E) and internode (IN), which indicates that they may also be involved in other biological functions in addition to flowering.

4.6. Study of the diurnal expression patterns of *BtLHY, BtTOC1, BtZTL and BtGI* genes in selected flowering and non-flowering tissues

Diurnal expression analyses of the four *B. tulda* genes (*BtLHY*, *BtTOC1*, *BtZTL* and *BtGI*) were conducted at four time points in a single day under both short day (SD) and long day (LD) conditions to understand their diurnal regulation in YLF and YLN. Diurnal expression level of *BtTOC1* and *BtGI* revealed the maximum expression in the afternoon, while *BtLHY* and *BtZTL* were in the morning under SD and LD conditions (**Figure 7**). Additionally, it was also observed that the *BtZTL* demonstrated a unimodal (morning) type of expression under LD while it became bimodal (morning and evening) under SD condition. Such divergences in expression may be important for regulation of these genes and their role in flowering.

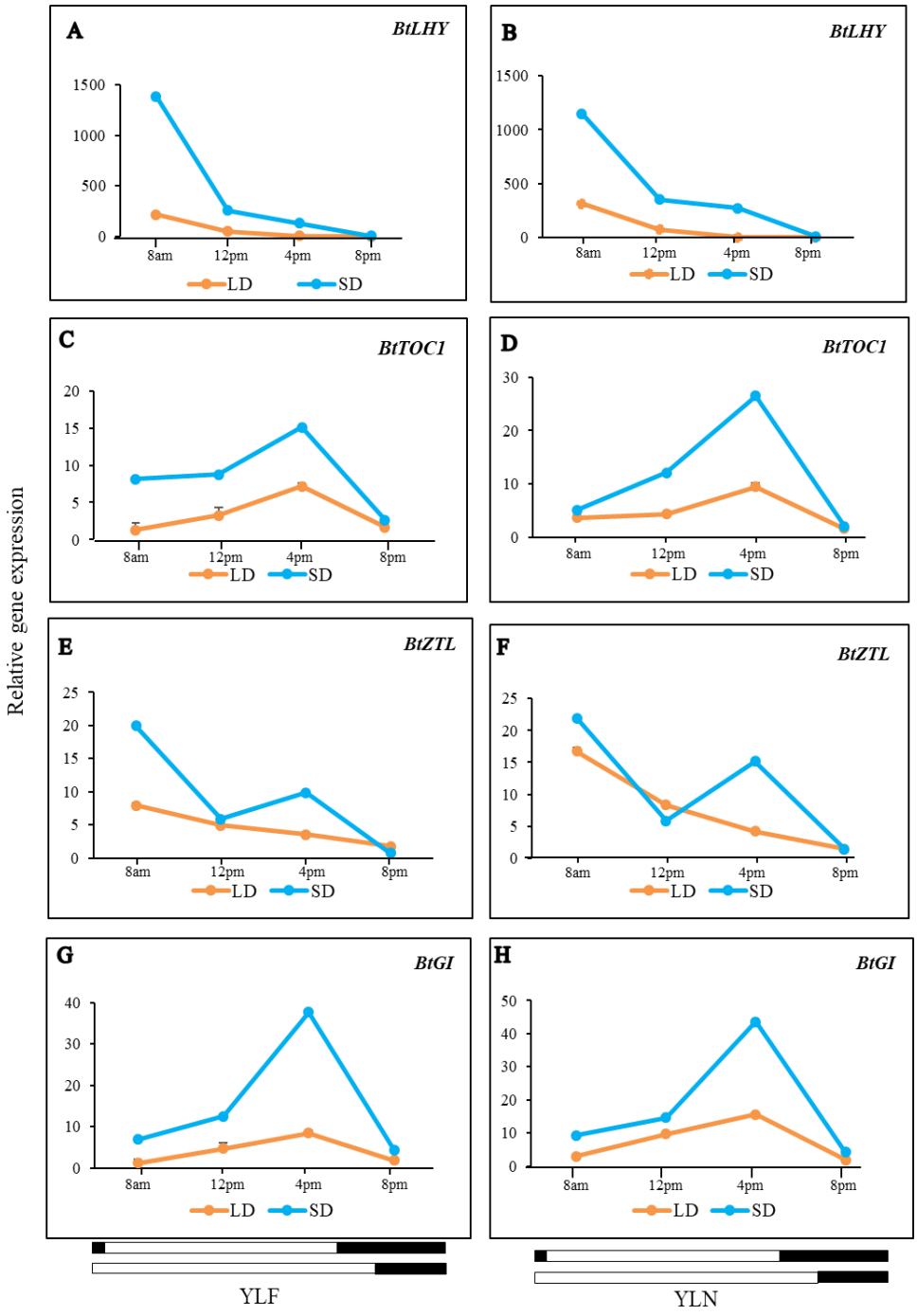

Figure 7. Analysis of diurnal expression pattern of circadian clock genes in *B. tulda*. The *eIF4α* was used as reference gene and rhizome as the calibrator. The $2^{-\Delta\Delta CT}$ method was used to measure the relative fold of the targeted genes. The fold change value of the three biological replicates ±SE were represented in each time point in LD and one biological replicate for SD. Abbreviations used: YLF- young leaf from flowering culm, YLN- young leaf from non-flowering culm.

4.7. *In silico* identification and comparisons of *cis*-acting elements present in the putative promoter regions of *LHY, TOC1, ZTL* and *GI* genes

To get further insight on gene expression divergence, 1500 bp upstream sequences of the four circadian clock genes (*LHY*, *TOC1*, *ZTL* and *GI*) identified from *P. heterocycla* were compared to that of related species. *In silico* search of *cis* acting elements demonstrated that the upstream regions of the identified *LHY*, *TOC1*, *ZTL* and *GI* genes were mainly composed of the light responsive elements such as CCAAT-box, G-box, GT1-motif, I-box, MBS, MYB recognition site, Sp1-motif, chs-Unit 1 m1-motif, ATCT-motif, GTGGC-motif, GA-motif, GATA- motif, AE-box, TCCC-motif, Box-4, BOX-II, LAMP element, CIRCADEAN element, 3AF-1-motif, CAG-motif along with hormonal/growth/stress responsive elements and the core promoter elements such as TATA box and CAAT box (**Figure 8**). Among all the monocotyledonous species least number of light responsive elements were recorded in *P. heterocycla LHY, GI* and *ZTL* genes. genes.

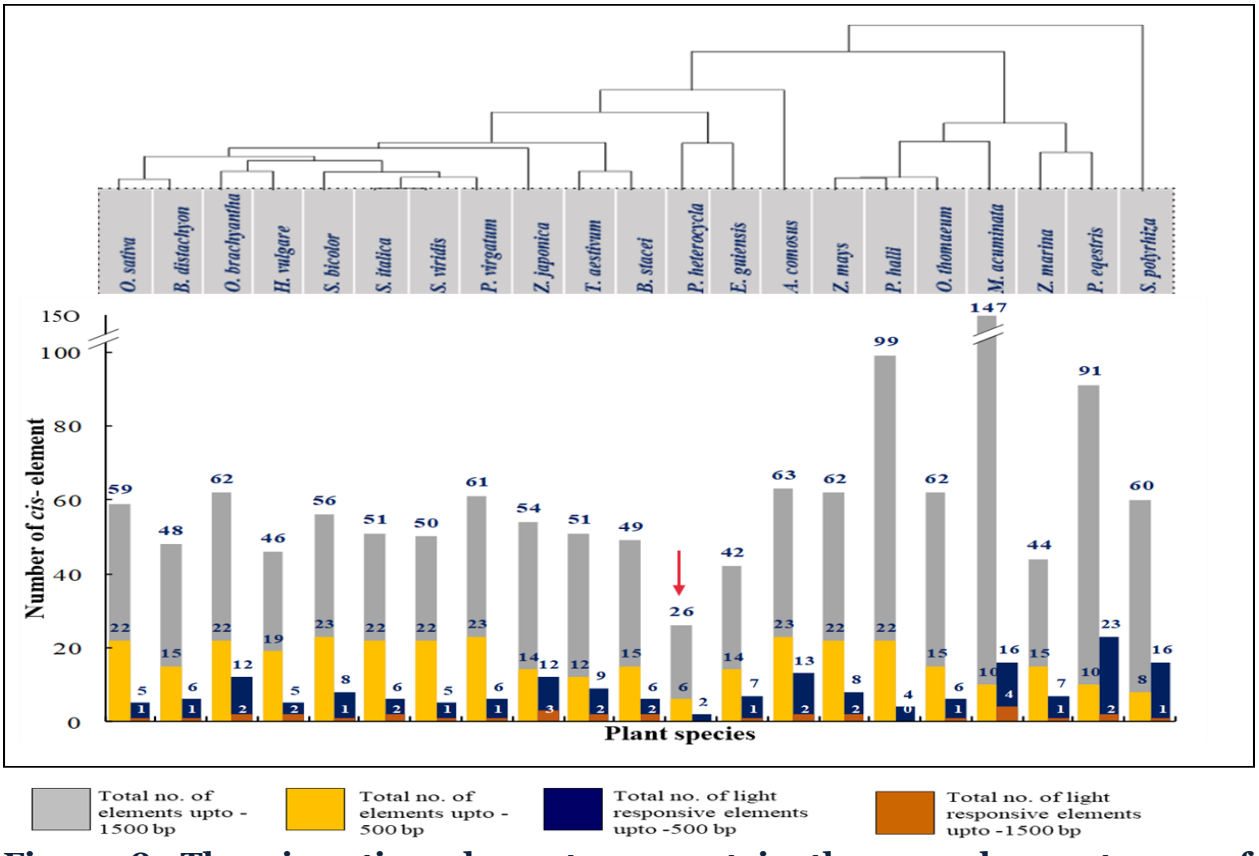

Figure 8. The *cis* acting elements present in the 1500 bp upstream of translation initiation site (ATG) of the identified *LHY* homologs

Objective 2: Identification, sequencing and detailed molecular characterization of important circadian clock integrator gene, *CONSTANS* in *B. tulda* Roxb.

4.8. *In silico* identification of *CONSTANS LIKE* genes from all the Poaceae plants having sequenced genomes

Among the four major clusters of *COL*s (group-I, II, III and IV), only a few members of group-I COLs are known to be involved in flowering. Therefore, in order to identify the homologous sequences of group-I *COL* genes from the available Poaceae species, BLASTP analysis was performed. Majority of the group-I *COL* homologs were found except group-Ib (homolog of *OsCOB*, **Table 5**). Interestingly, all the seven *COL* homologs were found in the sequenced bamboo genome *P. heterocycla* with high sequence homology with that of *O. sativa*. For example, for *OsCOA* 63%, for *OsCOB* 72%, for *OsCOC* 84%, for *OsCOD* 76.69%, for *OsCOE* 76.55%, for *OsCOF* 58.33% and for *OsCOG* 55.83% sequence homologies were recorded. Therefore, this sequence-based homology search analyses suggested that the group-I *COL*s were conserved in the temperate, perennial bamboo *P. heterocycla.*

Table 5. *In silico* identification of *CO* gene homologs identified from Poaceae. BLASTP analyses were performed using *O. sativa* amino acid sequences as quarries. Criteria used for BLAST analyses and subsequent selection of homologs were: identities (≥40%), E values (e^{-10}) and query coverage (≥40%). When multiple BLAST hits were obtained, only top BLASTP hits were considered for further analyses.

Plant species	*COA* homologs	*COB* homologs	*COC* homologs	*COD* homologs	*COE* homologs	*COF* homologs	*COG* homologs
O. brachyantha	OB06G19770	OB09G10860	OB04G24680	OB02G30990	OB06G30990	OB06G30990	OB08G17300
Z. mays	GRMZM2G 405368	NHF	ZMAC233888.1	GRMZM2G021777	GRMZM2G038783	GRMZM2G148772	GRMZM2G176173
H. vulgare	AAM74062.1	HVU0041G1317	HVU0042G1510	HVU0042G1510	HVU0042G1075	HVU0042G1075	HVU0045G0178
S. bicolour	SOBIC.010G115800.1	NHF	SOBIC.006G135100.1	SOBIC.004G211200.1	SOBIC.010G214000	SOBIC.004G063200.1	SOBIC.007G092000.1
T. aestivum	TAE17754G001	TAE04569G001	TAE35301G001	TAE11809G001	TAE33231G001	TAE10764G001	TAE01242G001
B. distachyon	BRADI1G43670.1	NHF	BRADI5G14600.2	BRADI3G48447.1	BRADI1G31280.1	BRADI3G05800	BRADI3G19010.1
B. stacei	BRAST07G100700.1	NHF	BRAST09G131600.1	BRAST04G141500.1	NHF	BRAST04G279200.1	NHF
S. itallica	SEITA.4G122700.1	NHF	SEITA.7G153500.1	SEITA.1G2 28800.1	SEITA.4G192300.1	SEITA.1G065300.1	SEITA.6G096800.1

S. viridis	SEVIR.4G194900.1	NHF	SEVIR.7G161700.1	SEVIR.1G2 32600.1	SEVIR.4G202100.1	SEVIR.1G064400.1	SEVIR.6G104600.1
P. hallii	PAHAL.D01761.1	NHF	PAHAL.G0 0773.1	PAHAL.A0 2488.1	PAHAL.D00662.1	PAHAL.A00517.1	NHF
P. virgatum	PAVIR.DA01464.1	NHF	PAVIR.J38013.1	PAVIR.AA 01385.1	PAVIR.DB00540.1	PAVIR.AA03360.1	PAVIR.FA01306.1
O. thomaeum	OROPETIUM_20150105_16080A	NHF	OROPETIUM_20150105_12006A	OROPETIUM_20150105_02724A	OROPETIUM_20150105_15716A	OROPETIUM_20150105_15716A	OROPETIUM_20150105_13880A
Z. japonica	ZJN_SC00131.1.G00960.1	NHF	ZJN_SC0000 4.1.G03470.1	NHF	ZJN_SC00008.1.G03620.1	NHF	NHF
P. heterocycla	PH01005551G0030	PH0100004 8G0270	PH0100078 0G0510	PH0100045 9G0310	PH01000481G0620	PH01000481G0620	PH01000831G0230

4.9. Isolation and molecular characterization of *CO* homologs from *B. tulda*

Two copies of *CO* homologs have been isolated from *B. tulda* and were submitted to NCBI (*BtCOA*- KY249523 and *BtCOB*- MF983714). Sequence based homology search analysis performed in NCBI database revealed that *BtCOA* demonstrated highest homology with the homolog of perennial tropical grass *Oryza rufipogon* (AFK31610.1, 85.08%), while *BtCOB* with the annual monocot crop *H. vulgare* (AAM74066.1, 76%).

4.10 *In silico* prediction of gene structures of *BtCOA* and *BtCOB* and comparison with homologs from related monocotyledonous species

The gene structure prediction analysis revealed that both the *BtCO* homologs were consisted of two exons and one intron (**Figure 9**), which is also true for all the CO homologs identified in other monocot species expect the *P. heterocycla COA*. Lengths of the identified *COL* homologs varied greatly from 704 bp (*P. heterocycla COG*) to 16033 bp (*O. sativa COB*). However, lengths of the CDS were quite conserved across *COL*s. For instance, the predicted lengths of *BtCOA* and *BtCOB* CDS were 1149 bp and 984 bp, respectively. In addition, all the identified *COL*s possessed two 43 amino acids long, B-boxes present in the N-terminal region and a 70 amino acids long, CCT domain located in the C-terminal region (**Figure 9**). The C-terminal part of the B-box 1 and N-terminal part of the B-box 2 were truncated in case of *BtCOB* (**Figure 10**),

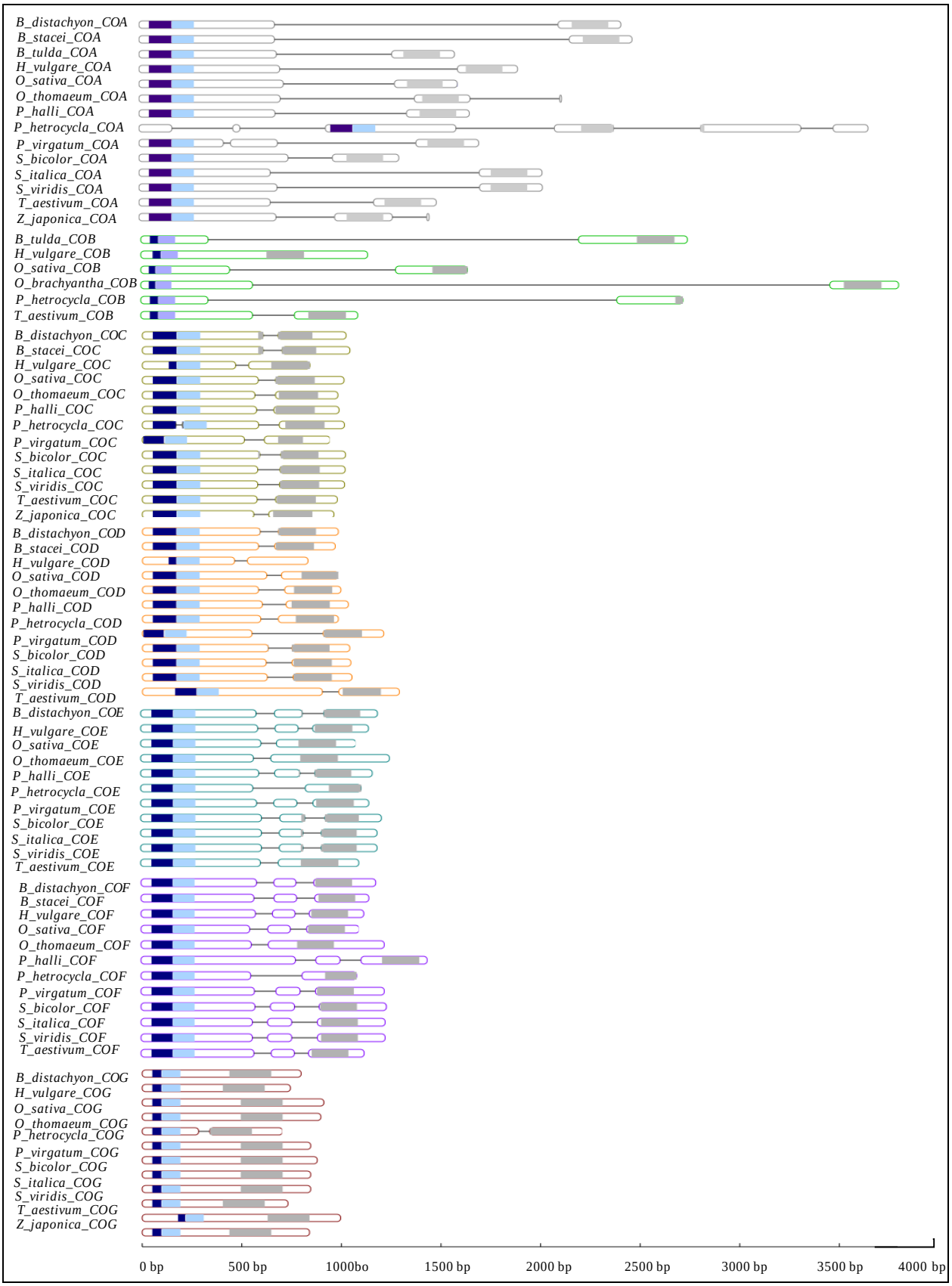
B_distachyon_COA
B_stacei_COA
B_tulda_COA
H_vulgare_COA
O_sativa_COA
O_thomaeum_COA
P_halli_COA
P_hetrocycla_COA
P_virgatum_COA
S_bicolor_COA
S_italica_COA
S_viridis_COA
T_aestivum_COA
Z_japonica_COA
B_tulda_COB
H_vulgare_COB
O_sativa_COB
O_brachyantha_COB
P_hetrocycla_COB
T_aestivum_COB
B_distachyon_COC
B_stacei_COC
H_vulgare_COC
O_sativa_COC
O_thomaeum_COC
P_halli_COC
P_hetrocycla_COC
P_virgatum_COC
S_bicolor_COC
S_italica_COC
S_viridis_COC
T_aestivum_COC
Z_japonica_COC
B_distachyon_COD
B_stacei_COD
H_vulgare_COD
O_sativa_COD
O_thomaeum_COD
P_halli_COD
P_hetrocycla_COD
P_virgatum_COD
S_bicolor_COD
S_italica_COD
S_viridis_COD
T_aestivum_COD
B_distachyon_COE
H_vulgare_COE
O_sativa_COE
O_thomaeum_COE
P_halli_COE
P_hetrocycla_COE
P_virgatum_COE
S_bicolor_COE
S_italica_COE
S_viridis_COE
T_aestivum_COE
B_distachyon_COF
B_stacei_COF
H_vulgare_COF
O_sativa_COF
O_thomaeum_COF
P_halli_COF
P_hetrocycla_COF
P_virgatum_COF
S_bicolor_COF
S_italica_COF
S_viridis_COF
T_aestivum_COF
B_distachyon_COG
H_vulgare_COG
O_sativa_COG
O_thomaeum_COG
P_hetrocycla_COG
P_virgatum_COG
S_bicolor_COG
S_italica_COG
S_viridis_COG
T_aestivum_COG
Z_japonica_COG
0 bp
500 bp
1000bo
1500 bp
2000 bp
2500 bp
3000 bp
3500 bp
4000 bp

Figure 9. Predicted models of *COL* genes identified in monocotyledonous plants. Exons were marked as hollow cylinders having conserved B-boxes and CCT domains and introns as solid lines. The B-box 1, B-box 2 and CCT domains were marked in dark blue, light blue and grey solid boxes, respectively.

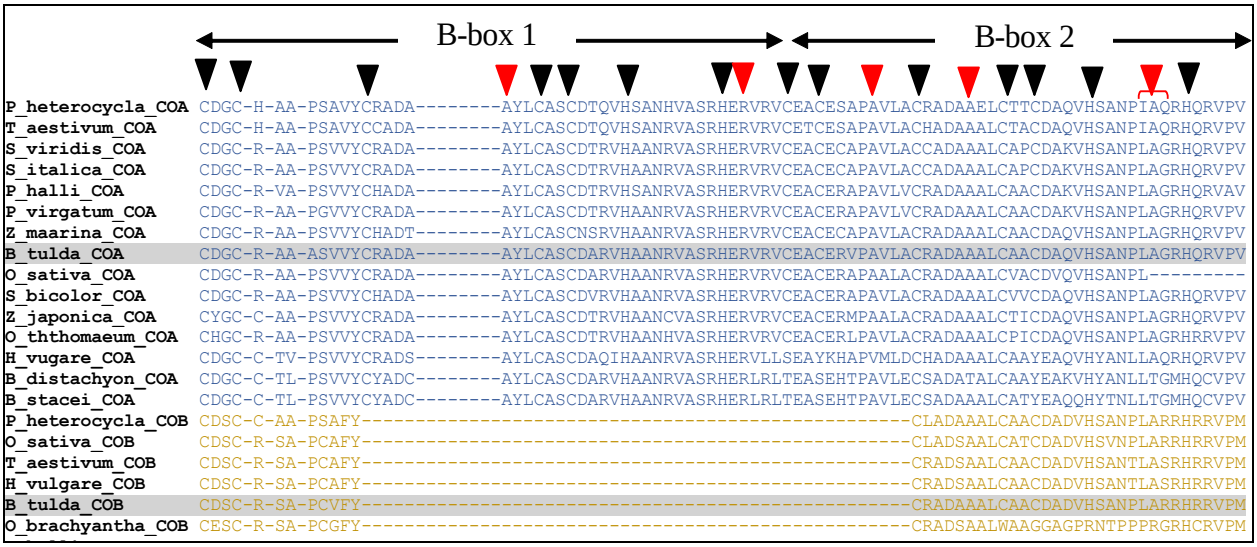

Figure 10. Comparison of B-box domains identified in the *CO* homologs obtained from *B. tulda* and comparison with related monocotyledonous species. The homologs identified from *B. tulda* were highlighted in grey. The characteristic C and H residues of B-box domains were indicated in black triangle. Amino acids marked as red triangle were functionally important as evidenced by mutational analyses.

while they were conserved in all the identified *COB*s and *COG*s. In plants, the CCT domain interacts with other DNA binding proteins such as *HAP3* and *HAP5* through nine conserved amino acids. Detailed sequence based analysis revealed that all these amino acids were conserved in *BtCOA*, while Arg33 was changed to Gln33 in *BtCOB*.

4.11. Phylogenetic relationships of *BtCOA* and *BtCOB* with homologous *COL*s identified from related monocotyledonous species

To get further insight into the evolutionary relationship of *BtCOA* and *BtCOB* genes, phylogetic tree was constructed using amino acid sequences (**Figure 11**). It was grouped into five major clusters. Clusters 1, 2 and 3 were comprised of members of *COA*, *COB*, *COG* homologs, cluster 4 was composed of the *COC* and *COD* homologs and cluster 5 possessed all the *COE* and *COF* homologs. In addition, the two *CO* homologs identified from *B. tulda* were split in two separate clusters. The *B. tulda COA* homolog was grouped into sub-cluster 1, while the homolog of

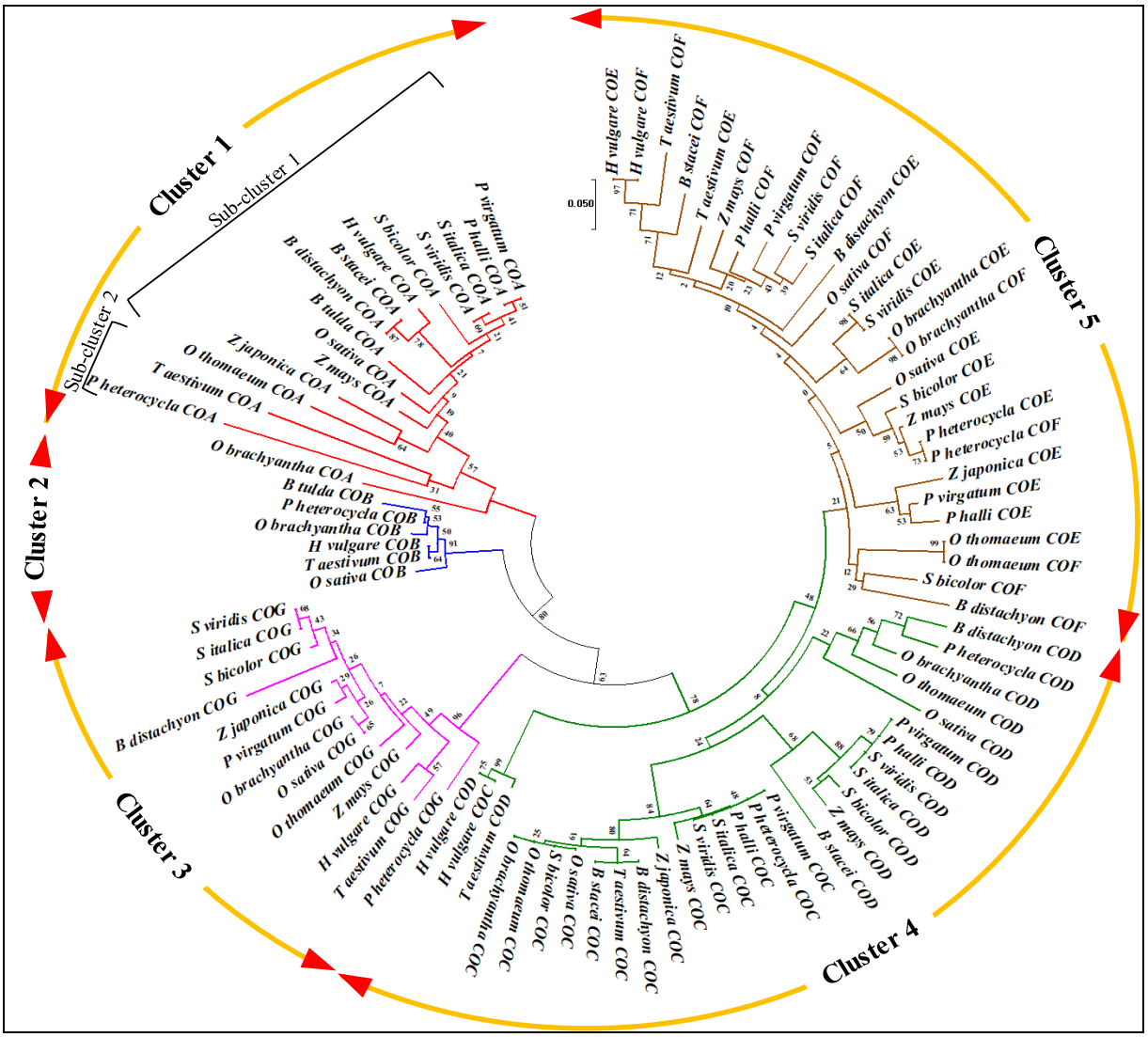

Figure 11. Phylogenetic analysis and gene model prediction of *COL* genes identified in monocots. Neighbour Joining (NJ) tree of the *CO* homologs was constructed with the full-length protein sequences by Mega 7.0 in default parameters using bootstrap value 1000.

temperate bamboo (*P. heterocycla*) was located in sub-cluster 2 within cluster 1 indicating sequence divergence between two homologous bamboo genes. In contrast, the *COB* homolog of *B. tulda* was clustered with that of temperate bamboo species *P. heterocycla* in cluster 2 (**Figure 11**).

4.12. Study of gene expression patterns of *BtCO* homologs in different flowering and non-flowering tissues

Transcriptional abundance of the two *BtCO* homologs was measured in six non-flowering and three flowering tissues. In the case of *BtCOA* and *BtCOB* genes, maximum level of expression

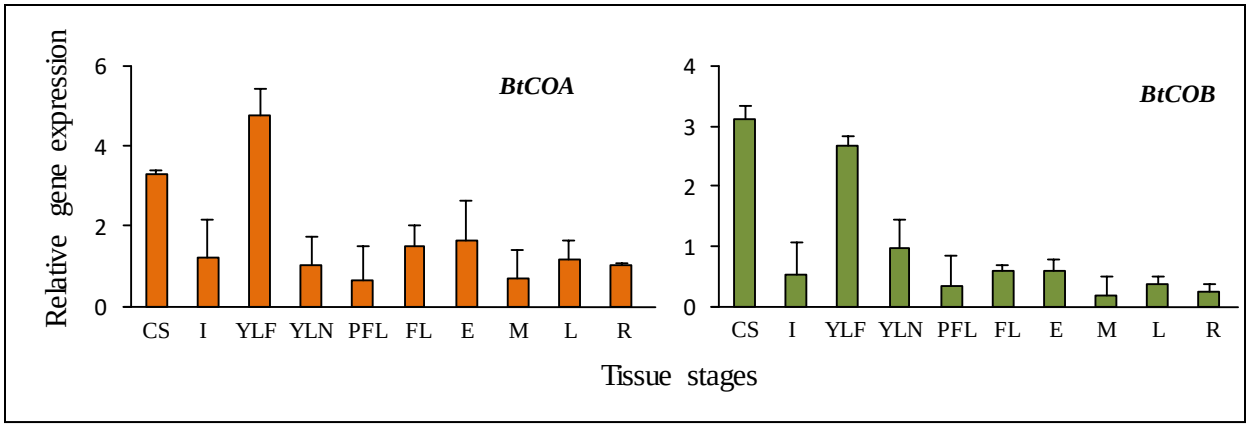

Figure 12. Gene expression analysis of circadian clock genes in *B. tulda*. The *eIF4α* was used as reference gene and rhizome as the calibrator. The $2^{-\Delta\Delta CT}$ method was used to measure the relative fold of the targeted genes. The fold change value of the three biological replicates ±SE were represented in each bar. Abbreviations used: CS: culm sheath, YLF- young leaf from flowering culm, I: internode, PFL: possible flag leaf, FL: flag leaf, E: early staged inflorescence bud, YLN- young leaf from non-flowering culm, M: middle staged inflorescence bud, L: late staged inflorescence bud, R: root.

was perceived in YLF and culm sheath (CS). The expression level was consistently low in all other selected tissues for both the genes. Interestingly, transcriptional accumulation was relatively high in YLF than CS for *BtCOA*, while such difference could not be observed for *BtCOB* (**Figure 12**). Therefore, this renders an important clue regarding the involvement of *BtCOA* in floral induction, as YLF is primarily connected with the flowering while CS is mostly vegetative.

4.13. Study of the diurnal expression of *BtCOA* and *BtCOB* in selected flowering and non-flowering tissues

Even though, spatial gene expression analysis of genes can provide some clues regarding gene functionality, the diurnal gene expression patterns of *BtCOA* and *BtCOB* genes were studied here to get further insight on the impact of circadian clock in regulating flowering. Interestingly, the diurnal expression pattern of *BtCOA* attained a peak in the afternoon and the suddenly decreased under both in SD and LD conditions. In contrast, the expression maxima of *BtCOB* was obtained in the morning (**Figure 13**). Therefore, the differential expression patterns of *BtCO* homologs in SD and LD condition suggested that *BtCOA* was possibly involved in flowering.

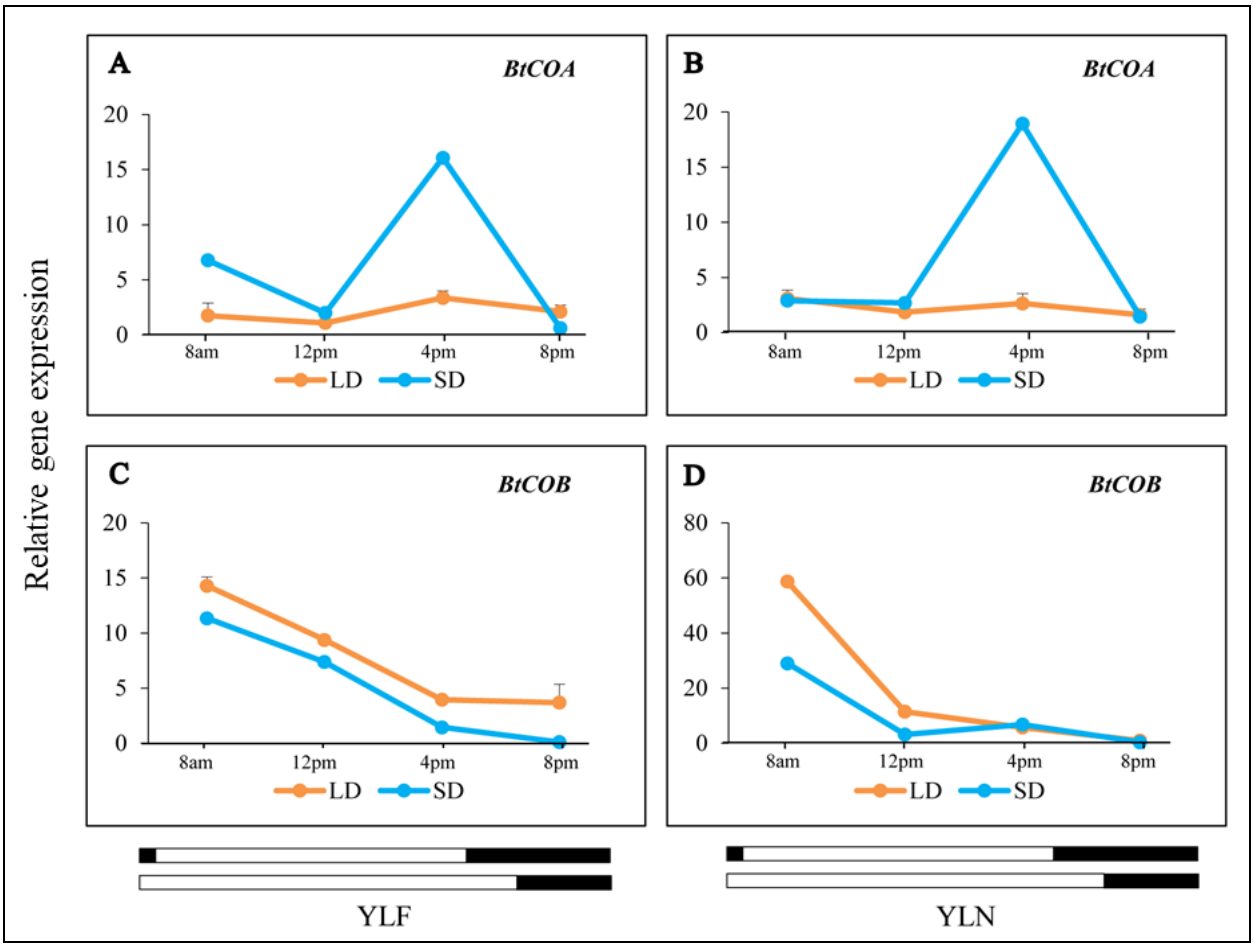

Figure 13. Analysis of diurnal expression pattern of circadian clock genes in *B. tulda*. The *eIF4α* was used as reference gene and rhizome as the calibrator. The 2-ΔΔCT method was used to measure the relative fold of the targeted genes. The fold change value of the three biological replicates ±SE were represented in each time point in LD and one biological replicate for SD. Abbreviations used: YLF- young leaf from flowering culm, YLN- young leaf from non-flowering culm.

4.14. *In silico* identification and comparison of *cis*-acting elements present in the *COA* and *COB* homologs

One of the major factors behind the divergence of gene expression pattern is divergence of promoter sequences. Therefore, *in silico* study was conducted to identify *cis* acting elements present in the 1500 bp upstream region of the *CO* homologs. This analysis revealed that the upstream region of the two *CO* homologs was comprised of 15 light responsive elements and 15-17 hormonal/growth/stress responsive elements along with the core promoter elements (**Figure 14**). The number of light responsive elements present in the *COA* homologs ranged from 4 (*P. heterocycla COA*) to 18 (*S. bicolor COA*), while it was 8 (*P. heterocycla COB*) to 22 (*O. sativa*

COB) for *COB* (**Figure 14**). Interestingly, least number of light responsive elements were observed in the *COA* and *COB* genes of *P. heterocycla*.

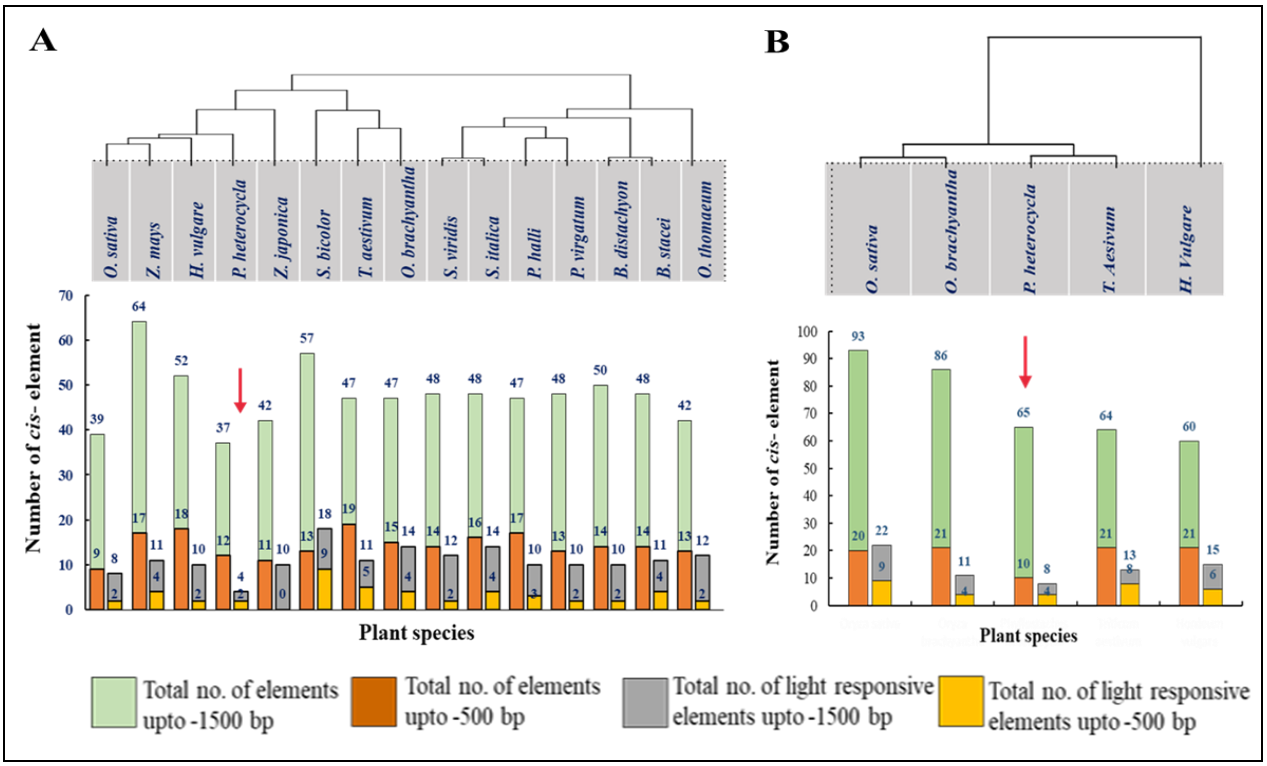

Figure 14. The *cis* acting elements identified upto 1500 bp upstream region from translation initiation site (ATG). A. *COA* homologs. **B.** *COB* homologs.

Objective 3: Identification and transgene mediated functional characterization of BtFD1 and BtFD2 gene homologs of B. tulda Roxb.

4.15. Identification of *FD1* and *FD2* gene homologs in the sequenced Poaceae and non-Poaceae genomes

At least one homolog of *OsFD1* and *OsFD2* could be found in all the monocot crops such as *Z. mays, H. vulgare, S. bicolour, T. aestivum* and also in the annual and perennial grasses *B. distachyon, B. stacei, S. itallica, S. viridis, O. thomaeum, Z. japonica, D. oligosanthes, P. hallii, P. virgatum*. Homologs of *OsFD1* and *OsFD2* were also identified in the temperate (e.g., *S. veitchii, P. heterocycla*) and tropical bamboo (e.g., *B. tulda*). All the identified bamboo *FD1* genes possessed high sequence homology with that of *O. sativa*. Interestingly, in our set BLASTP criteria, no homologous sequences of *OsFD2* could be identified in any of the non-Poaceae species studied (**Table 6**). Therefore, it can be hypothesized that *FD2* genes are conserved in Poaceae, but not in the non-Poaceae clade.

Table 6. *In silico* identification of *FD* gene homologs identified from the Poaceae and non-Poaceae memebers of monocotyledonous plants. BLASTP analyses was performed using *O. sativa* amino acid sequences as queries. Criteria used for BLAST analyses and subsequent selection of homologs were: identities (≥40%), E values (e^{-10}) and query coverage (≥40%).

Monocot plant groups	Plant species	Locus ID/ Accession Number of best BLAST hit	
		***FD1* homologs**	***FD2* homologs**
Poaceae	*Oryza sativa*	OS09G36910	OS06G50830
	Oryza brachyantha	OB09G24570	OB02G45490
	Zea mays	ZM00001D022613	ZM00001D036392
	Hordeum vulgare	HVU0041G3471	HVU0045G2520
	Sorghum bicolour	SOBIC002G280800	SOBIC010G269000
	Triticum aestivum	TAE32871G001	TAE36408G001
	Aegilops tauschii	XP020151150	XP020177215
	Bracypodium distachyon	BRADI4G36587	BRADI1G29920
	Bracypodium stacei	BRAST05G197700	BRAST07G238100
	Setaria itallica	SEITA2G291300	SEITA4G282900
	Setaria viridis	SEVIR2G302300.1	SEVIR4G295100
	Panicum hallii	PAHALB03671.1	PAHALD00176
	Panicum virgatum	PAVIRJ04490.1	PAVIRJ04759
	Oropetium thomaeum	OROPETIUM_ 20150105_06520	OROPETIUM_ 20150105_16409
	Zoysia japonica	ZJNSC00122.1. G00160.1	ZJNSC00008.1. G00500.1
	Dicanthelium oligosanthes	OEL23045	OEL37294
	Bambusa tulda	MF983712	MH142577
	Phyllostachys heterocycla	PH01000511G0500	PH01001986G0070
	Sasa veitchii	BAS04368	NHF
Non-Poaceae	*Zostera marina*	ZOSMA70G00700	NHF
	Spirodela polyrhiza	SPIPO3G0017700	NHF
	Musa acuminata	MAC12G0975	NHF
	Ananas comosus	ACO009346	NHF
	Phalenopsis equestris	PEQU26966	NHF
	Dendrobium catenatum	XP020692523.1	NHF
	Elaeis guineensis	EGU0206G0533	NHF
	Phoenix dactylifera	XP008780307	NHF

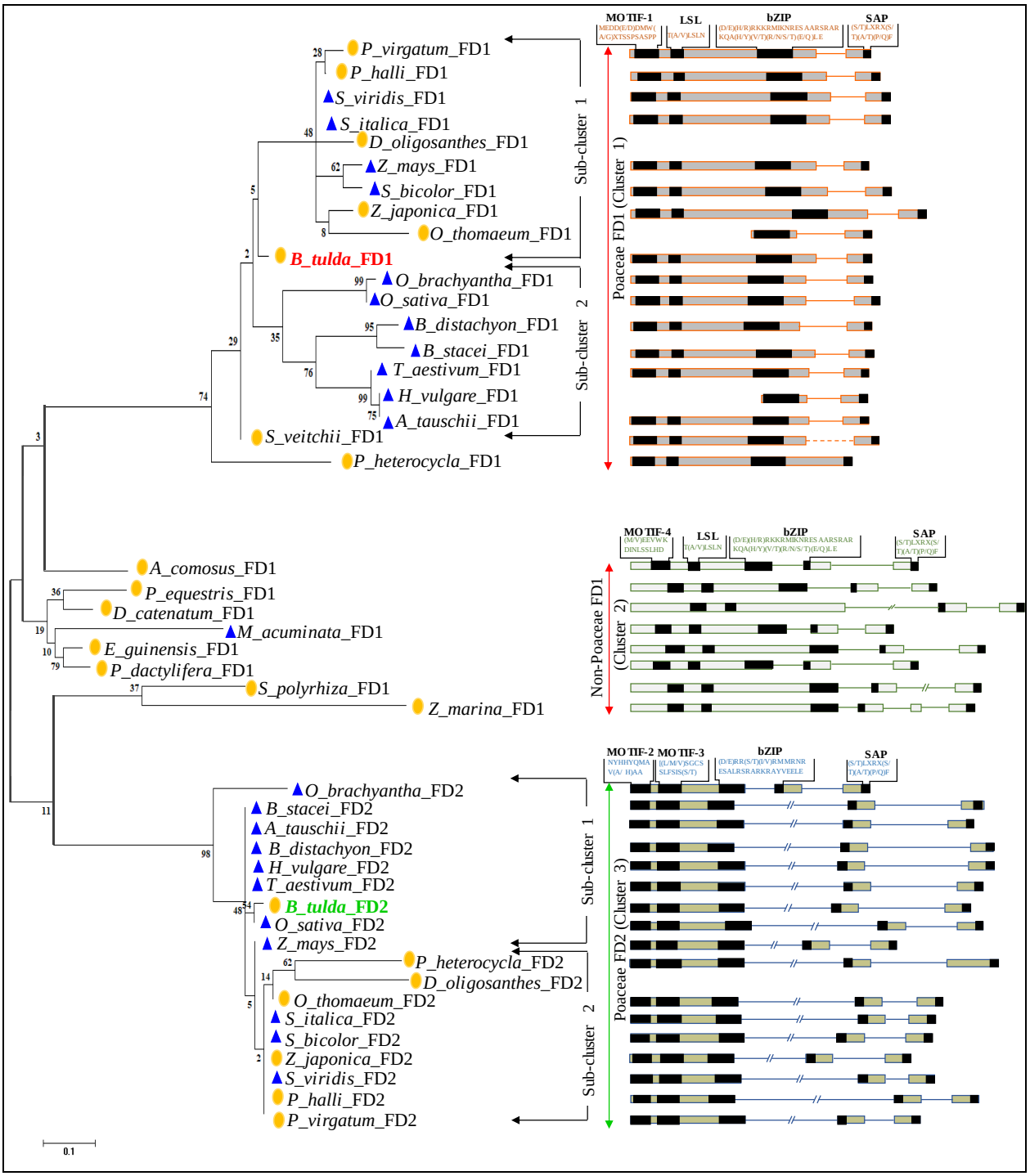

Figure 15. Phylogenetic analysis and gene model prediction of four *FD* genes identified in monocots. The Neighbour Joining (NJ) tree of the *FD* homologs was constructed with the full-length protein sequences by Mega 7.0 in default parameters using 1000 bootstrap value. In the gene structures exons were marked as rectangles having conserved motifs marked in solid black boxes and introns as solid lines. Annual species were marked in blue tringle and perennials in yellow circles.

4.16. Phylogenetic analysis and comparison of *BtFD1* and *BtFD2* gene structure with their homologs present in the Poaceae and non-Poaceae plants

In phylogenetic analysis the *FD* genes were broadly divided into three major clusters. The cluster 1 was comprised of *FD1* homologs identified from the Poaceae species that included crops, annual and perennial grasses and three bamboo species (*B. tulda, S. veitchii, P. heterocycla*). The cluster 2 was comprised of *FD1* homologs from all the non-Poaceae members and the cluster 3 was comprised of all *FD2* homologs (**Figure. 15**). The cluster specific for Poaceae *FD1* i.e. cluster 1 was sub divided into two major sub-clusters. Sub-cluster 1 hosted *FD1* sequences obtained from annual (*Z. mays, S. bicolour, S. itallica, S. viridis*) and perennial (*O. thomaeum, Z. japonica, D. oligosanthes, P. hallii, P. virgatum*) plants**,** while sub-cluster 2 hosted rest of the annuals. The members of the three bamboo *FD1*s were located in Cluster 1. Although, *B. tulda FD1* was located in sub-cluster 1, FD1 genes of P. *heterocycla* and *S. veitchii* were out-clustered (**Figure. 15**). Cluster 3 was also subdivided into two sub-clusters like that of cluster 1. Here *B. tulda FD2* along with all annual plants were placed in sub-cluster 1 while *P. heterocycla FD2* was observed in sub-cluster 2 with many members from annual and perennial plant groups.

Gene structure analyses revealed that the Poaceae specific *FD1* genes contains two exons and a single intron except *Phyllostachys heterocycla* (**Figure. 15**). All the Poaceae specific *FD1* genes possessed their identifying motif-1 [MEDD(E/D)DMW(A/G)XTSSPSASPP], LSL [T(A/V)LSLN], bZIP and SAP [(S/T)LXRX(S/T)(A/T)(P/Q)F] domains. The *FD2* genes possessed three exons and two introns and contains the identifying motif-2 [NYHHYQMAV(A/H)AA], motif 3 [(L/M/V)SGCSSL FSIS(S/T)], bZIP and modified SAP.

4.17. Gene expression analyses of *BtFD* genes across five different seasons

To get further insight into the involvement of *BtFD1* and *BtFD2* in controlling flowering, their expression in young leaves were studied in five different seasons: summer, monsoon, autumn, winter and spring. Transcriptional expression level of *BtFD1* transcript remained higher than *BtFD2* all throughout the year. However, the expression level of *BtFD1* in YLF was highest in winter, which was the season just before the onset of sprouting (**Figure. 16**). The sudden rise of expression level of *BtFD1* in autumn followed by it's expression maxima in the winter and overlap with the flower induction in winter suggested that there might be a strong link between

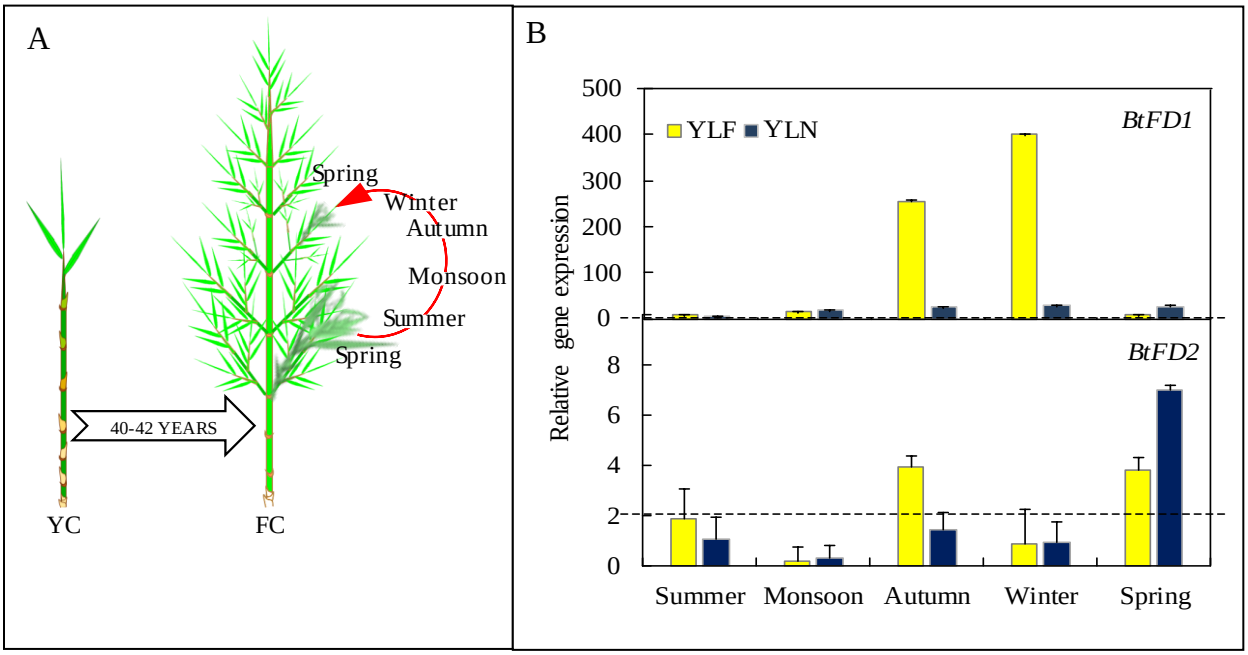

Figure 16. Study of relative gene expression levels of *BtFD* homologs in five different seasons.
A. Schematic representation of flowering cycle observed in case of sporadic flowering of *B. tulda*. **B.** Relative gene expression levels of *BtFD1* and *BtFD2* in young *B. tulda* leaves collected in five different seasons. The *eIF4α* was used as the reference gene and rhizome as the calibrator. The 2-ΔΔCT method was used to measure the relative fold of the targeted genes. The fold change value of the three biological replicates ±SE were represented in each

season induced transcriptional accumulation of *FD* and flowering. In contrary, *BtFD2* expression was barely detectable and quite comparable in YLF and YLN among all of the seasons. Therefore, the differential rates of expression of the both *BtFD1* and *BtFD2* throughout the year suggested a highly variable gene expression correlation in order to seasonal changes on flowering.

4.18. *In silico* interaction study of *BtFD* gene homologs with the promoter of *AP1*

Docked structures of both *BtFD1* and *BtFD2* bZIP models predicted positive interactions with the CRE DNA (TGACGT). Several amino acid residues at the basic region, spanning from His133 to Gln153 in *BtFD1* and Arg87 to Arg107 in *BtFD2* were found to interact with TGACGTCA consensus sequence present in the 21 bp CRE DNA sequence (**Figure. 17**). Therefore, it can be concluded that both *BtFD1* and *BtFD2* bZIP domains can interact with *AP1* promoter region.

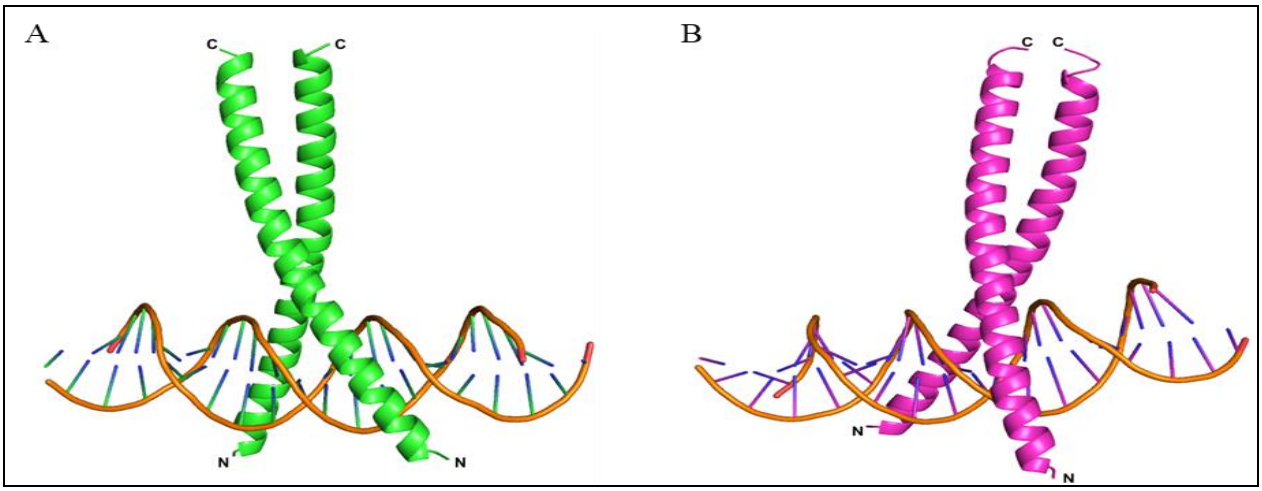

Figure 17. Model structures of bZIP domains. A. BtFD1, **B.** BtFD2 shows overall ability to bind with CRE DNA

4.19. Study of phenotypic characters of *35Sp::BtFD1* and *35Sp::BtFD2* transgenic plants

Since both *BtFD1* and *BtFD2* were capable to interact with C-box element present in the promoter of *AP1*, the transgenic approach was carried forward to investigate the role of *BtFD* homologs. Phenotypes of *35Sp::BtFD1* plants revealed either extremely delayed flowering or nonflowering (**Figure. 18**), while flowering time was delayed in case of *35Sp::BtFD2* plants

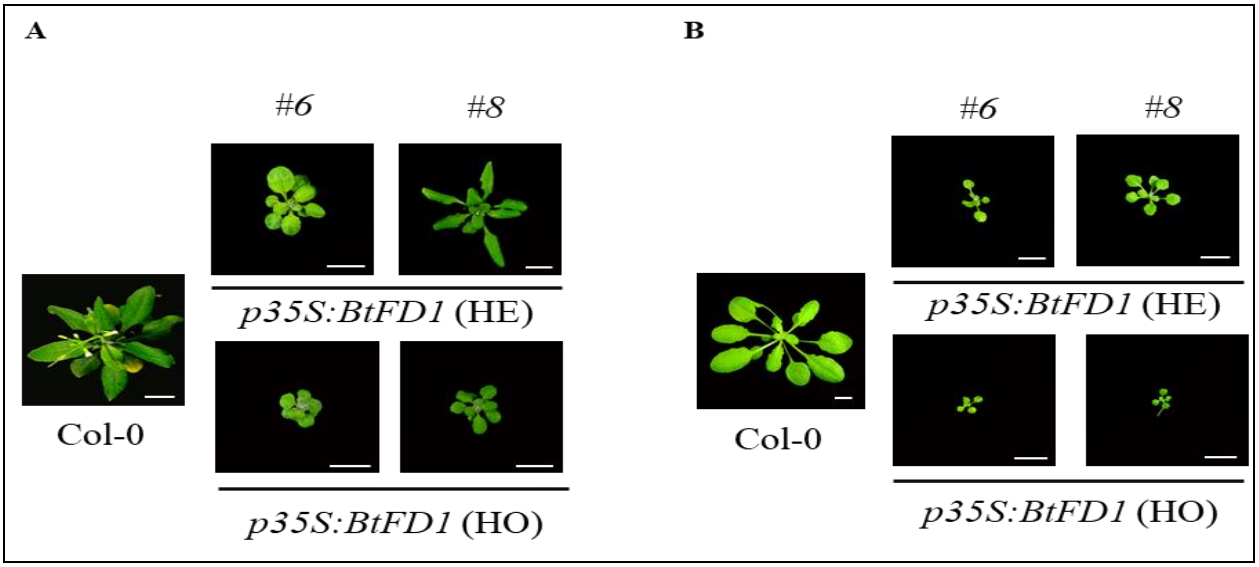

Figure 18. Phenotypes of transgenic plants overexpressing *BtFD1* under Cauliflower Mosaic Virus (CaMV) 35S promoter in *A. thaliana* Col-0 plants. A. Plants grown under LD condition (16-h light and 8-h dark) for four weeks. **B.** Plants grown under SD condition (10-h light and 14-h dark) for four weeks. The scale bar represents 1 cm. The abbreviations used: HO- homozygous, HE- heterozygous.

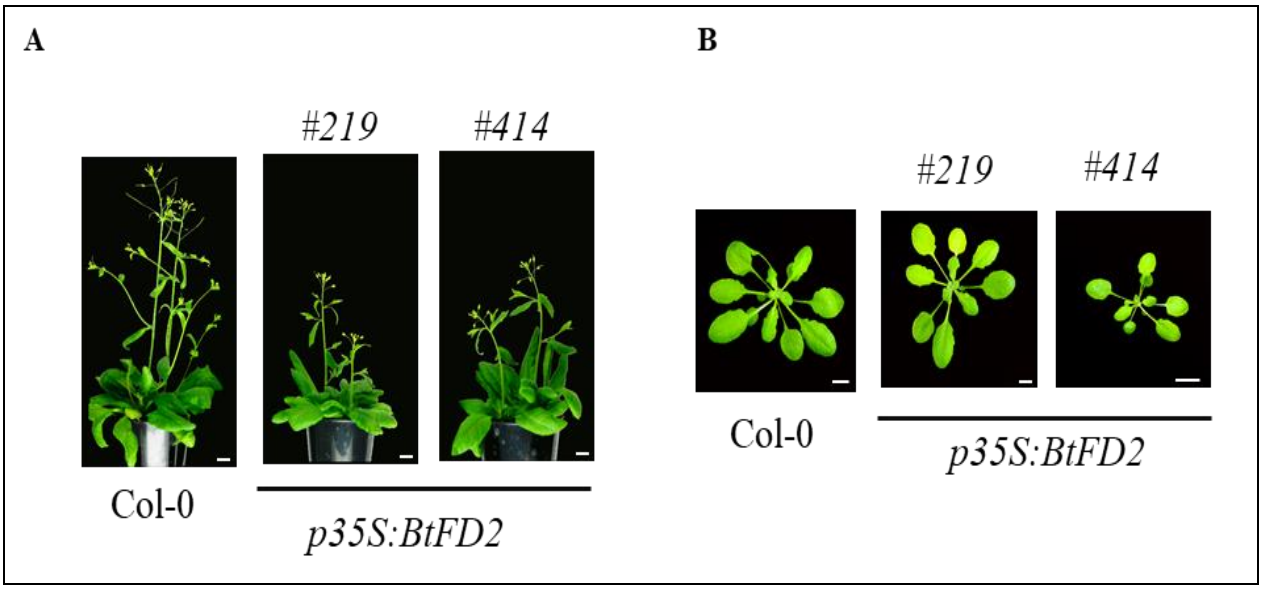

Figure 19. Phenotypes of transgenic plants overexpressing *BtFD2* under Cauliflower Mosaic Virus (CaMV) 35S promoter in *A. thaliana* Col-0 plants. **A.** Plants grown under LD condition (16-h light and 8-h dark) for four weeks. **B.** Plants grown under SD condition (10-h light and 14-h dark) for four weeks. The scale bar represents 1 cm.

(**Figure. 19**). Additionally, the vegetative growth was also surprisingly reduced in *35Sp::BtFD1* plants, while it was not much affected for the *35Sp::BtFD2* plants (**Figure. 18, 19**). Apart from the number of leaves, the size of leaves were drastically reduced in the *35Sp::BtFD1* plants. However, for *35Sp::BtFD2* plants this was quite comparable to the wild type. In addition, it was also found that under SD condition the vegetative growth was severely affected in *35Sp::BtFD1* homozygous lines compared to hemizygous plants (**Figure. 18**). Therefore, these results suggest that *BtFD1* clearly had some effect on both vegetative as well as reproductive growth.

5. Discussion

5.1. Important circadian clock genes are present and possibly involved in bamboo flowering

The circadian rhythmicity with respect to flowering has extensively been studied in the reference plants *A. thaliana* and *O. sativa* along with a few annual crops (Murakami et al. 2007; McClung, 2008). However, no studies have been performed to characterize the circadian clock genes and their relation to flowering in the perennial monocotyledonous plant bamboo. Therefore, in this study one of the main challenges was identification of the important circadian gene homologs

from temperate as well as tropical bamboos. In this study BLASTP analysis indicated that the circadian genes (*LHY, TOC1, ZTL* and *GI*) were present in the temperate bamboo *P. heterocycla*. In addition, at least one copy of these homologs was obtained from *B. tulda* and they possessed high sequence similarity with other monocotyledonous homologous genes certainly at the biologically active domain regions. Further from the evolutionary point of view this study suggested that *BtLHY, BtTOC1* and *BtGI* genes were phylogenetically closer to the temperate bamboo *P. heterocycla* which was not the case for *BtZTL*. This observation revealed that evolution as well as divergence in sequences of the clock genes may have happened after the speciation events. In addition, the detailed spatio temporal gene expression analysis of the circadian clock genes (*BtLHY, BtTOC1* and *BtGI*) in bamboo suggested that they possibly play important role in flowering.

5.2. The two *CO* homologs isolated from *B. tulda* possibly play diverse role with respect to flowering

CONSTANS (*CO*) is the circadian clock integrator gene which accepts ultimate light signal coming from the clock genes and upregulate *FT* for flower induction (Redei, 1962). In addition to the reference plant *A. thaliana* (Putterill et al., 1995), extensive studies have already been conducted in 27 dicotyledonous plant species to understand the role of *CO* in flowering. In contrast, detailed characterization of *CO* gene family in the monocotyledonous plant group was restricted to only a few species such as *H. vulgare, T. aestivum, L. perenne* and *F. pratensis*. Therefore, in this study *CONSTANS LIKE* genes had been identified from 15 Poaceae species which includes several crops as well as annual and perennial grasses along with the only sequenced temperate bamboo *P. heterocycla*. Moreover, two *CO* homologs (*BtCOA, BtCOB*) were identified from the tropical bamboo *B. tulda. a*. Interestingly, phylogenetic clustering revealed that the *BtCOA* was found quite distant from that of temperate bamboo *P. heterocycla* and more related to other annual as well as perennial species including *O. sativa* while *BtCOB* was close to that of *P. heterocycla*. Detailed expression analyses of *BtCOA* and *BtCOB* genes in different tissues revealed that the transcripts of *BtCOA* was abundantly present in flowering tissues (YLF) while *BtCOB* was in the vegetative tissues. Also the diurnal expression of *BtCOA* revealed a characteristic peak at the end of the day while it was observed in morning for *BtCOB*. Taken together, all of these findings conclude that *BtCOA* may primarily be associated with flowering, but not *BtCOB*.

5.3. *BtFD1* is involved in vegetative growth other than flowering

Consequent upon activation, *FT* interacts with *FD* in the cytoplasmic region of the shoot apical meristem, for the induction of flowering (Abe et al., 2005; Wigge et al., 2005). Other than transition to flowering, *FD* plays crucial roles in inflorescence development in pea (Sussmilch et al., 2015). In perennial poplar and vine kiwifruit *FD* homologs are involved in alternative growth cessation and flowering (Parmentier-Line and Coleman et al., 2016; Varkonyi-Gasic et al., 2013). All these findings suggest that the function of *FD* gene clade is not only restricted to flower induction but also in growth and development of inflorescence. In this study, detailed sequence characterization and gene expression analysis suggested that *BtFD1* is more likely associated with flowering. The *BtFD1* overexpressing *A. thaliana* plants exhibited extreme reduction in vegetative growth with respect to number as well as size of the rosettes suggesting their possible involvement in controlling the vegetative growth of the plants. In addition to the suppression of the vegetative growth the *35Sp::BtFD1* revealed either extremely delayed flowering or nonflowering phenotype. This may be an indirect effect of extreme vegetative growth retardation. It was also reported in *O. sativa* that overexpression of *AtFD* and *AtFDP* resulted in down-regulation of the cell wall growth responsive genes such as *EXTENSIN, EXPANSIN, XTH1* (Jang et al., 2017). This hypothesis was supported since the expression level of floral meristem identity genes such as *MADS14/ 15/ AP1* remained high in the *FD* overexpressed *O. sativa* transgenic lines (Jang et al., 2017).

6. Literatures cited

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T (2005) FD, a bZIP protein mediating signals from the floral pathway integrator *FT* at the shoot apex. Science 309, 1052-6.

Alabadı´ D, Oyama T, Yanovsky MJ, Harmon FG, Ma´s P, Kay SA (2001) Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. Science 293, 880-883.

Ben-Naim O, Eshed R, Parnis A, Teper-Bamnolker P, Shalit A, Coupland G, Samach A, Lifschitz E (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. The Plant Journal 46, 462–476.

Blackman BK (2017) Changing Responses to Changing Seasons: Natural Variation in the Plasticity of Flowering Time. Plant Physiology 173, 16–26.

Boss PK, Bastow RM, Mylne JS, Dean C (2004) Multiple pathways in the decision to flower: enabling, promoting, and resetting. The Plant Cell. 16: S18-S31.

Das M, Haberer G, Panda A, Das Laha S, Ghosh TC, Schäffner AR (2016) Expression pattern similarities support the prediction of orthologs retaining common functions after gene duplication events. Plant Physiology 171, 2343–57.

Das M, Reichman JR, Haberer G, Welzl G, Aceituno FF, Mader MT, Watrud LS, Pfleeger TG, Gutiérrez RA, Schäffner AR, Olszyk DM (2010) A composite transcriptional signature differentiates responses towards closely related herbicides in *Arabidopsis thaliana* and *Brassica napus*. Plant Molecular Biology 72, 545-56.

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of *CONSTANS*-Like gene families in barley, rice, and *Arabidopsis*. Plant Physiology 131, 1855–1867.

Herrero E, Kolmos E, Bujdoso N, Yuan Y, Wang M, Berns MC, Uhlworm H, Coupland G, Saini R, Jaskolski M, Webb A, Goncalves J, Davis SJ (2012) *EARLY FLOWERING4* recruitment of

EARLY FLOWERING3 in the nucleus sustains the *Arabidopsis* circadian clock. Plant Cell 24, 428−443.

Ito S, Niwa Y, Nakamichi N, Kawamura H, Yamashino T, Mizuno T (2008) Insight into missing genetic links between two evening-expressed pseudo-response regulator genes TOC1 and PRR5 in the circadian clock-controlled circuitry in *Arabidopsis thaliana*. Plant Cell Physiology 49, 201–213.

Jang S, Li HY, Kuo ML (2017) Ectopic expression of *Arabidopsis FD* and *FD PARALOGUE* in rice results in dwarfism with size reduction of spikelets. Scientific Reports 7, 44477.

Janzen DH (1976). Why bamboos wait so long to flower. Annual Review in Ecology and Systematics 7, 347–391.
John CK Nadgauda RS (2002) Bamboo flowering and famine. Current Science 82 (3), 261–262.

Kim WY, Fujiwara S, Suh SS, Kim, J, Kim Y, Han L, et al. (2007) ZEITLUPE is a circadian photo receptor stabilized by GIGANTEA in blue light. Nature 449, 356–360.

Li CX, Dubcovsky J (2008) Wheat FT protein regulates VRN1 transcription through interactions with FDL2. Plant Jornal 55, 543-554.

Lou P, Wu J, Cheng F, Cressman LG, Wang X, McClung CR (2012) Preferential retention of circadian clock genes during diploidization following whole genome triplication in *Brassica rapa*. Plant Cell 24, 2415–2426

McClung CR (2008).Comes a time. Current Opinion in Plant Biology 11, 514–520.

Murakami M, Tago Y, Yamashino T, Mizuno T. 2007. Comparative overviews of clock-associated genes of Arabidopsis thaliana and Oryza sativa. Plant & Cell Physiology 48, 110–121.

Parmentier-Line CM and Coleman GD (2016) Constitutive expression of the Poplar FD-like basic leucine zipper transcription factor alters growth and bud development. Plant Biotechnology Journal 14, 260–270.

Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80, 847–857.
Qin Z, Cai, X, Huang J (1989) Seed characteristics and natural regeneration of arrow bamboo (*Bashania fangiana*). Journal of Bamboo Resource 8, 1–12.

Redei GP (1962) Super vital mutants of *Arabidopsis*. Genetics 47, 443–460.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Teisher JK, Clark LG, Barberá P, Gillespie LJ, Zuloaga FO (2017) A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. Journal of Systemics and Evolution 55, 259-290

Takata, N., Saito, S., Saito, C. T., Nanjo, T., Shinohara, K., and Uemura, M. (2009). Molecular phylogeny and expression of poplar circadian clock genes, LHY1 and LHY2. New Phytol. 181, 808–819.

Taylor AH, Reid DG, Qin Z, Hu J (1991) Spatial patterns and environmental associates of bamboo (*Bashania fangiana*) after massflowering in southwestern China. Bulletin of Torrey Botanical Club 118, 247–254.
Tian X (1987) Ecological observation on flowering and regeneration of bamboo forests in Mt. Qinling. Journal of Bamboo Resource 2, 39–43.
Tremblay R, Colasanti J (2006) Floral Induction. Annual Plant Reviews. 20, 28-46.
Trujillo D (2007) 'Bamboo structures in Colombia'. The Structural Engineer, March 2007, 25-30

Tsuji H, Nakamura H, Taoka K, Shimamoto K (2013) Functional Diversification of FD Transcription Factors in Rice, Components of Florigen Activation Complexes. Plant and Cell Physiology 54, 385-397.

Varkonyi-Gasic E, Moss SMA, Voogd,C, Wang T, Putterill J, Hellens RP (2013) Homologs of FT, CEN and FD respond to developmental and environmental signals affecting growth and flowering in the perennial vine kiwifruit. New Phytologist 198, 732746.

Wenkel S. Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, Coupland G (2006) *CONSTANS* and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. The Plant Cell 18, 2971–2984.

Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. Science 309, 1056–1059.